

Universidad de San Carlos de Guatemala Facultad de Ingeniería Escuela de Ingeniería en Ciencias y Sistemas

IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALÍTICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS

Herbert Rafael Reyes Portillo

Asesorado por Ing. Marlon Francisco Orellana López

Guatemala, junio 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALÍTICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA FACULTAD DE INGENIERÍA POR

HERBERT RAFAEL REYES PORTILLO ASESORADO POR ING. MARLON FRANCISCO ORELLANA LÓPEZ

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, JUNIO 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martinez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz Gonzélez
SECRETARIA	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. César Augusto Fernéndez Céceres
EXAMINADOR	Ing. Herman Igor Véliz Linares
EXAMINADOR	Ing. Miguel Ángel Cancinos Rendón
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALÍTICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas con fecha 1 de noviembre de 2021.

Herbert Rafael Reyes Portillo



Ciudad de Guatemala, 04 de mayo de 2022

Ing. Carlos Azurdia Coordinador Trabajos de Tesis Escuela de Ingeniería en Ciencias y Sistemas Facultad de Ingeniería

Estimado Ing. Azurdia:

Por este medio informo que en mi papel de asesor he revisado el trabajo de Tesis titulado: "IMPLEMENTACION DE HERRAMIENTA PARA ANALITICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS", elaborado por el estudiante Herbert Rafael Reyes Portillo quién se identifica con DPI 2993840550101 y registro académico 201612114.

Con base a la evaluación realizada hago constar que este trabajo de graduación cuenta con mi aprobación.

Atentamente,

"Id y Enseñad a todos"

Ing. Marlon Francisco Orellana López Asesor de Tesis



Universidad San Carlos de Guatemala Facultad de Ingeniería Escuela de Ingeniería en Ciencias y Sistemas

Guatemala 9 de mayo de 2022

Ingeniero Carlos Gustavo Alonzo Director de la Escuela de Ingeniería En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante HERBERT RAFAEL REYES PORTILLO con carné 201612114 y CUI 2993 84055 0101 titulado "IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALITICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS" y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo aprobado.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,





Ing. Carlos Alfredo Azurdia Coordinador de Privados y Revisión de Trabajos de Graduación



LNG.DIRECTOR.114.EICCSS.2022

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador de área y la aprobación del área de lingüística del trabajo de graduación titulado: IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALÍTICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS, presentado por: Herbert Rafael Reyes Portillo, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería.

"ID Y ENSEÑAD A TODOS"

Msc. Ing. Carlos Gustavo Alonzo Director Escuela de Ingeniería en Ciencias y Sistemas



Guatemala, mayo de 2022

Ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, -Escuela de Ciencias, Regional de Ingeniería Sanitaria y Recursos Hidráulicos (ERIS), Maestría en Sistemas Mención construcción y Mención Ingeniería Vial. Carreras: Ingeniería Mecánica, Ingeniería Electrónica, Ingeniería en Ciencias y Sistemas, Licenciatura en Matemática, Licenciatura en Física. Centros: de Estudios Superiores de Energía y Minas (CESEM). Guatemala, Ciudad Universitaria, Zona 12, Guatemala, Centroamérica.



Decanato Facultad de Ingeniería 24189101-24189102 secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.402.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: IMPLEMENTACIÓN DE HERRAMIENTA PARA ANALITICA DE DATOS SOBRE EL PIB DE GUATEMALA UTILIZANDO LA CIENCIA DE DATOS, presentado por: Herbert Rafael Reyes Portillo, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

SHVERSIDAD DE SAN CARLOS DE GUATEMAL DECANA FACULTAD DE INGENIERÍA Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, junio de 2022

AACE/gaoc

Escuelas: Ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, - Escuela de Ciencias, Regional de Ingeniería Sanitaria y Recursos Hidráulicos (ERIS). Post-Grado Maestría en Sistemas Mención Ingeniería Vial. Carreras: Ingeniería Mecánica, Ingeniería Electrónica, Ingeniería en Ciencias y Sistemas. Licenciatura en Matemática. Licenciatura en Física. Centro de Estudios Superiores de Energía y Minas (CESEM). Guatemala, Ciudad Universitaria, Zona 12. Guatemala, Centroamérica.

ACTO QUE DEDICO A:

Dios	Por guiarme durante toda la carrera y darme las
	ruerzas para nunca darme por vencido.
Mis padres	Iris Portillo de Reyes y Rafael Reyes. Ser mi ejemplo para seguir y principal apoyo, ya que sin
	ellos no hubiera sido posible lograrlo.
Mis hermanos	Roberto, André y Ruby Reyes, por estar ahí.
Mis amigos	Ruth Lechuga, Kevin Mejia, Emely Garcia y Javier Golón, por brindarme su apoyo incondicional en todo momento.
Ingeniero	Marlon Orellana, por su apoyo incondicional, ser una guía en este proceso para que fuera posible lograrlo.

AGRADECIMIENTOS A:

Universidad de San	Por ser el alma máter donde pude incrementar
Carlos de Guatemala	mis conocimientos y darme los recursos para crecer profesionalmente.
Facultad de Ingeniería	Por el conocimiento que adquiri en sus instalaciones.
Ingenieros	Miguel Cancinos, Sharonn Pú, Glenda García, a todos aquellos que enseñan por vocación y dejaron una huella en mi proceso educativo.
Amigos de la Facultad	Carlos Hernández, Sergio Solís, Javier Lima, Yaiza Pineda, Jeackelin Montenegro y Jeannira Sic por los buenos momentos durante el proceso.

ÍNDICE GENERAL

ÍNIDIC					
INDIC					
LISTA DE SÍMBOLOS					
GLOSARIO					XI
RESL	JMEN				XIII
OBJE	TIVOS				XV
0202					
1.	ECONO	MÍA EN GI	JATEMALA		
	1.1.	Contexto	histórico		
	12	Banco de	Guatemala	(BANGUAT)	Δ
	1.2.				
	1.3.	Actividade	es economic	as de Guatem	ala
	1.4.	Producto	Interno Brut	0	
		1.4.1.	Cálculo de	PIB	9
			1.4.1.1.	Según la proc	ducción o el valor añadido. 9
			1.4.1.2.	Según el flujo	de gastos9
			1.4.1.3.	Según los ing	resos o flujo de rentas 10
2	CIENCI	A DE DATO	os		11
	0.1				11
	Ζ.Ι.	Ciencia d	e dalos		
		2.1.1.	Ciclo de vio	da de la Cienci	a de Datos 11
	2.2.	Analítica	de Datos		
		2.2.1.	Definición o	de análisis de o	datos 12
			2.2.1.1.	Tipos de anál	isis de datos 13
				2.2.1.1.1.	Análisis descriptivo 13
				2.2.1.1.2.	Análisis de diagnóstico. 13

2.2.1.1.3. Análisis predictivo...... 14

				2.2.1.1.4.	Análisis prescriptivo	14
		2.2.2.	Integració	n de datos		14
			2.2.2.1.	Herramientas	s y técnicas de Integracio	ón
				de datos		15
	2.3.	Lenguaj	e R			16
		2.3.1.	Entorno R			16
		2.3.2.	RStudio			17
	2.4.	Análisis	de tendencia	as a nivel de ar	alítica de datos	17
		2.4.1.	Método N	aive		18
		2.4.2.	Método E	xponential Smo	oothing	18
		2.4.3.	BATS Y T	BATS		19
			2.4.3.1.	TBATS		19
			2.4.3.2.	BATS		19
		2.4.4.	Modelos A	ARIMA / SARIM	1A	20
			2.4.4.1.	ARIMA		20
			2.4.4.2.	SARIMA		20
3.	ANALÍ	FICA Y VIS	SUALIZACIO	N DE DATOS		21
	3.1.	Microsof	ft Power Bi			21
		3.1.1.	Tecnologí	as de Microsof	t Power Bi	22
			3.1.1.1.	Power Query	/	22
			3.1.1.2.	DAX		23
4.	HERRA	MIENTA F	PARA EL AN	IÁLISIS DE PR	ONÓSTICOS	25
	4.1.	ETL de o	datos			25
		4.1.1.	Extracciór	n de datos		25
			4.1.1.1.	Descarga de	la fuente de datos de	
				Banguat		25

			4.1.1.2.	Vinculación de	e la data con Microsoft	
				Power Bi		26
		4.1.2.	Transforma	ción de datos.		28
			4.1.2.1.	Limpieza, filtra	ado, transformación e	
				inclusión de la	a fuente de datos	28
				4.1.2.1.1.	Transformación	
					automática por medio de	Э
					Power Query	29
			4.1.2.2.	Sumarización	de resultados	31
5.	ANALÍTI	CA DE DA	TOS			35
	5.1.	Análisis e	stadístico de	escriptivo		35
		5.1.1.	Sumario de	estadística de	escriptiva total país	35
		5.1.2.	Sumario de	estadística de	escriptiva por categoría	36
		5.1.3.	Análisis de	e datos explo	tatorios y análisis de	
			<i>outliers</i> p	oor categoría		38
	5.2.	Análisis p	redictivo			62
		5.2.1.	Análisis pre	edictivo del tota	Il del país	62
		5.2.2.	Análisis pre	edictivo por cate	egoría	64
6.	CREACI	ÓN DE TA	BLERO DE	CONTROL		81
	6.1.	Generació	ón de visualiz	zaciones		81
	6.2.	Publicació	on de las visi	ualizaciones		85
CONC	CLUSION	ES				93
RECC	MENDA	CIONES				95
BIBLI	OGRAFÍA	۱				97
APÉN	IDICES				1	01
ANEX	0				1	11

IV

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	RStudio	17
2.	Partes de Microsoft Power Bi	22
3.	Fases de Power Query	23
4.	Microsoft Excel información PIB	
5.	Inicio de Microsoft Power Bi	27
6.	Carga de Microsoft Excel en Microsoft Power Bi	28
7.	Transformar con Power Query	29
8.	Power Query	30
9.	Reemplazar en Power Query	31
10.	Creación de Tabla en Microsoft Power Bi	32
11.	Creación de sumarización por DAX	32
12.	Tabla sumarizada	33
13.	EDA de comercio y reparación de vehículos	39
14.	Outlier de comercio y reparación de vehículos	39
15.	EDA de agricultura, ganadería, silvicultura y pesca	40
16.	Outlier de agricultura, ganadería, silvicultura y pesca	41
17.	EDA de explotación de minas y canteras	42
18.	Outlier de explotación de minas y canteras	42
19.	EDA de industrias manufactureras	43
20.	Outlier de industrias manufactureras	44
21.	EDA de suministro de electricidad, agua y saneamiento	45
22.	Outlier de suministro de electricidad, agua y saneamiento	45
23.	EDA de construcción	46

24.	Outlier de construcción	47
25.	EDA de transporte y almacenamiento	48
26.	Outlier de transporte y almacenamiento	48
27.	EDA de actividades de alojamiento y servicio de comidas	49
28.	Outlier de actividades de alojamiento y servicio de comidas	50
29.	EDA de información y comunicaciones	51
30.	Outlier de información y comunicaciones	51
31.	EDA de actividades financieras y de seguros	52
32.	Outlier de actividades financieras y de seguros	53
33.	EDA de actividades inmobiliarias	54
34.	Outlier de actividades inmobiliarias	54
35.	EDA de actividades profesionales, científicas y técnicas	55
36.	Outlier de actividades profesionales, científicas y técnicas	56
37.	EDA de actividades de servicios administrativos y de apoyo	57
38.	Outlier de Actividades de servicios administrativos y de apoyo	57
39.	EDA de administración pública y defensa	58
40.	Outlier de administración pública y defensa	59
41.	EDA de enseñanza	60
42.	O <i>utlier</i> de enseñanza	60
43.	EDA de salud	61
44.	O <i>utlier</i> de salud	62
45.	Análisis predictivo total país con Covid	63
46.	Análisis predictivo total país sin Covid	64
47.	Análisis predictivo actividades de alojamiento y de servicio	de
	comidas	65
48.	Análisis predictivo actividades de servicios administrativos y	de
	ароуо	66
49.	Análisis predictivo actividades financieras y de seguros	67
50.	Análisis predictivo actividades inmobiliarias	68

51.	Análisis predictivo actividades profesionales científicas y técnicas	69
52.	Análisis predictivo administración publica y defensa	70
53.	Análisis predictivo agricultura, ganadería, silvicultura y pesca	71
54.	Análisis predictivo comercio y reparación de vehículos	72
55.	Análisis predictivo de construcción	73
56.	Análisis predictivo de enseñanza	74
57.	Análisis predictivo de explotación de minas y canteras	75
58.	Análisis predictivo de industrias manufactureras	76
59.	Análisis predictivo de información y comunicaciones	77
60.	Análisis predictivo de salud	78
61.	Análisis predictivo de suministro de electricidad, agua y saneamiento	79
62.	Análisis predictivo de transporte y almacenamiento	80
63.	Dataset cargado en Microsoft Power Bi	81
64.	Forecasting with Arima	82
65.	Slicer de categoría	83
66.	Agregando gráfica de forecasting con ARIMA	84
67.	Visualizaciones en Microsoft Power Bi	84
68.	Aplicación de Microsoft Power Bi en Microsoft 365	85
69.	Creación de Espacio de trabajo.	86
70.	Publicación de visualización	87
71.	Tablero publicado en Web de Microsoft Power Bi	87
72.	Visualizacion de Tablero Publicado	88
73.	Creación de Microsoft SharePoint	89
74.	Publicar informe en la web	90
75.	Código para publicar en sitio web	90
76.	Insertar código en Microsoft SharePoint	91
77.	Pegando código de Microsoft Power Bi	92
78.	Dashboard publicado en Microsoft SharePoint	92

TABLAS

I.	Sumario estadístico total pais en millones de quetzales	.35
II.	Sumario estadístico por categoría en millones de quetzales	.36

LISTA DE SÍMBOLOS

Significado
Exploratory Data Analysis
Extraer, transformar, cargar
Producto Interno Bruto
Moneda quetzal

Х

GLOSARIO

- Almacén de datosEs un depósito de datos generados y recopilados por
los diversos sistemas
- **Científico de datos** Es un perfil profesional que traduce los grandes volúmenes de información disponibles provenientes de fuentes de información masivas.
- DatasetConjunto de datos estructurado compuesto por filas y
columnas que representan tanto registros como
variables.
- Inteligencia Artificial Programa de computación diseñado para realizar determinadas operaciones que se consideran propias de la inteligencia humana, como el autoaprendizaje.
- Machine Learning Es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente.
- MiddlewareEs software que se sitúa entre un sistema operativo ylas aplicaciones que se ejecutan en él.
- Variable Cuantitativa Variable estadística que puede expresarse a través de cifras.

XII

RESUMEN

Por medio de la ciencia de datos, se pueden realizar pronósticos en base a la ingesta de información que se le brinde, gracias a que día a día ha aumentado su uso por medio de los modelos estadísticos se obtienen resultados y en base a estos se puede realizar un análisis, o bien tomarlo como referencia de qué conducta puede tomar ese análisis y qué tanto énfasis debe darse en él. El uso de la ciencia de datos en la economía no está de más, ya que viendo el comportamiento de la economía podemos obtener un pronóstico que nos pueda ayudar con prevenciones o tomarlo como referencia de que inversiones tomar.

Gracias a la información histórica que se posee de BANGUAT, es posible la generación de una herramienta que realice pronósticos en base al top 5 de rubros en el PIB de Guatemala, para posteriormente realizar un análisis de ellos y dar una conclusión. Utilizar el top 5 de actividades del PIB de Guatemala y ver su comportamiento nos brinda esa información de torno a que gira el futuro del país, a cuáles actividades hay que tomarles mas importancia, y por que se debe el crecimiento de ciertas actividades.

Para la representación de la data generada por el pronóstico se utilizará un tablero de control que nos muestre la información de forma organizada y concisa. El análisis del PIB de Guatemala es un tema muy interesante para tocar con la ciencia de datos, ya que se ha caracterizado a Guatemala como un país netamente agrícola, pero se plantea por medio del pronóstico ver si realmente esta tendencia puede persistir en los próximos años o los demás sectores serán quienes realmente dominen la actividad económica del país.

XIII

XIV

OBJETIVOS

General

Generar una herramienta de analítica de datos para el análisis de actividades económicas del Producto Interno Bruto de Guatemala independientemente del periodo de evaluación.

Específicos

- Generar un análisis de los sectores económicos de Guatemala de los que se pronostique un mayor crecimiento y se les debe dar un mayor énfasis porque dada las oportunidades que puedan generar oportunidades.
- Hacer uso de las técnicas de ciencia de datos para explorar aquellas actividades económicas en crecimiento según el reporte "Producto Interno Trimestral" del Banguat.
- Generar un tablero de control (*dashboard* no público) que integre información histórica y actual para pronosticar tendencias de las actividades económicas en crecimiento.
- Analizar el impacto que tuvo la pandemia de Covid-19 con relación al producto bruto interno y como hará que evolucione en la economía del actual del país.

XVI
INTRODUCCIÓN

Por medio de una herramienta que utilice la ciencia de datos se pueden realizar pronósticos acerca del comportamiento que va a tomar cierta variable a lo largo del tiempo. Con la información brindada por estos pronósticos, se puede realizar un análisis acerca de las decisiones y/o consideraciones a tomar.

El Producto Bruto Interno determina la suma de bienes y servicios que posee un país, por lo que analizando cada categoría que conformaan a éste, da a conocer de mejor manera como se está desarrollando la economía dentro del país. Además, por medio de un análisis de PIB, también se pueden ver reflejados aspectos como la situación política, cultural y social afectan a las producciones de cada país.

Con la herramienta empleada para este trabajo de investigación, fue posible realizar un pronóstico en base a los datos brindados por el banco de Guatemala, y con éste ver la tendencia que tiene cada categoría en los próximos años. El análisis de cada tendencia orienta a cuáles categorías tienden a poseer un mayor crecimiento y cuáles se han visto reducidas.

La situación de COVID-19 se ha visto impactado en la economía del país, por las restricciones tomadas por gobierno y situación mundial, por lo que muchas categorías del PIB se vieron afectadas, pero al ser un evento extraordinario, afecta los pronósticos y por ello se realiza la comparativa de situación con covid-19 y sin éste.

1. ECONOMÍA EN GUATEMALA

1.1. Contexto histórico

En todo país, por medio del PIB (Producto Bruto Interno), término utilizado para determinar la suma de bienes y servicios de un país. Se ha tomado como referencia de como medir la riqueza de un país, ya que este indicador toma lo anteriormente mencionado. Además, este indicador denota si un país se encuentra en crecimiento o caída de su economía.

Cuando el PIB aumenta se tiene un incremento en el gasto e inversión de un país, y se ve reflejado en crecimiento de ingresos de las empresas, por lo que genera mejores salarios y oportunidades de empleo, y con este ciclo se incrementa la economía del país debido a que con mayores ingresos, más oportunidades de poder realizar compras y así inyectar recursos financieros a la economía; por lo contrario cuando este cae se disminuye el consumo, se tienen menores ingresos en las empresas, y la tasa de desempleo incrementa o el ingreso salarial de las personas disminuye, y la economía comienza a entrar en una recesión.

Guatemala principalmente ha sido conocido como un país netamente agrícola, al tomar en cuenta datos históricos del Banco en Guatemala¹, en el año de 1950 la Agricultura, silvicultura, caza y pesca representaba un 33,14 % del PIB del país, y ya para finales de la década el año 1969 representaba un 27,29 %, se puede observar que en el transcurso de 2 décadas, la representación de

¹ Banguat. *PIB Período 1950 a 1959. http://www.banguat.gob.gt/inc/ver. asp?id=estaeco/sr/sr019&e=46332. Consulta: 3 de octubre 2021.*

la agricultura en el PIB de Guatemala cayó un 5,85 puntos, es decir que con el paso del tiempo se fue teniendo un mayor desarrollo en otras actividades del país, y obteniendo una mayor representación en la estructura del producto interno bruto.

En el año de 1950, la segunda actividad económica más importante de Guatemala era Comercio al por mayor y al por menor, con una representación del 26,68 %, es una diferencia notable con respecto a la Agricultura, pero al ver el año 1969, esta actividad económica represento el 28,33 % del PIB del país, superando a la Agricultura. 30 años después en el año 2000, con un mayor PIB, el sector agrícola representaba un 22 % del total del PIB del país, en vez de tener una mayor representación para llamar a Guatemala un país (Netamente Agrícola) se puede observar que su representación va disminuyendo.

Finalmente viendo el resultado del primer trimestre del año 2021 según Banguat², la Agricultura tiene una representación del 10,65 %, cuando vemos la comparativa con el año 1950, el sector agrícola disminuyo una tercera parte, mientras que el sector que antiguamente era clasificado como "Comercio al por mayor y al por menor" actualmente tiene el nombre de Comercio y Reparación de Vehículos representa el 18,24 %, puede que sea un menor valor a comparación de los cincuentas, pero vemos que esta actividad se mantiene predominante en el PIB del país (Rank #1 año 2021).

² Banguat. *Producto Bruto Interno Trimestral.* https://banguat.gob.gt/sites/default/files/banguat/cuentasnac/PIB2013/PDF_graficas_y_cuadros _estadisticos.pdf. Consulta: octubre 2021.

El producto Interno Bruto de Guatemala, tuvo su mayor apogeo de crecimiento entre los años 1963 a 1977 según el Banco Mundial³, teniendo su mayor pico en 1963 con un 9,541 % de crecimiento y con excepción del año 1975 que fue del 1,95 % mientras que los otros años se mantuvieron con una media mayor al 5 %.

Pero luego de estos años de mayor crecimiento, comenzó una recesión en el país, el PIB fue en caída, llegando a su punto más bajo en 1982 cayendo hasta un -3,53 %, sin embargo, se fue recuperando en las siguientes décadas, pero sin llegar a superar el pico máximo de crecimiento de 1963, la cifra más alta ha sido del 6,338 % en el 2007. Los siguientes años ha sido un PIB de picos de altas y bajas, del 2017 al 2019 se estaba teniendo una tendencia de crecimiento luego de una caída en el año 2016, hasta que llego la pandemia del COVID-19.

En el 2020 la economía mundial fue impactada debido a la pandemia COVID-19, gran parte de los países terminaron el año con porcentajes negativos de crecimiento, a pesar de esta situación, el impacto en Guatemala no fue tan grande a comparación de otros países, pero tuvo una caída de 1,5 % versus el 2019. Durante este año de pandemia el top 5 de rubros que del PIB de la economía de Guatemala fueron: 1. Comercio y reparación de vehículos, 2. Industrias manufactureras, 3. Agricultura, ganadería, silvicultura y pesca, 4. Actividades Inmobiliarias y 5. Información y comunicaciones.

Podemos notar que, en la tendencia actual, la agricultura ya está en tercer lugar, ya no ocupa ni si quiera el segundo puesto, y algo curioso, que a pesar de que el rubro de Información y comunicaciones es relativamente nuevo (ya que

³ Banco Mundial. *Crecimiento del PIB (% anual) – Guatemala.* https://datos.bancomundial.org/indicator/NY.GDP.MKTP.KD.ZG?locations=GT. Consulta: 4 de octubre de 2021.

anteriormente no estaba en las actividades económicas del PIB del país) ahora forma parte del TOP 5 de actividades económicas.

Cabe destacar con el análisis del año 2020, se pueden observar los sectores que a pesar de un acontecimiento que cambio totalmente al mundo como la pandemia COVID-19, estos continúan representando las actividades más importantes del país. Con lo que, podemos suponer que, a pesar de una crisis económica mundial, podemos ver cuales sectores son los que brindan más oportunidades porque a pesar de la tragedia que conllevo el COVID-19, gracias a la transformación en el mundo, muchos sectores pudieron crecer y desarrollarse más, mientras que otros lamentablemente vivieron el efecto opuesto.

Sectores como la construcción, actividades de alojamiento y servicios de comida se vieron mayormente afectadas por las disposiciones que el gobierno manejo durante época de COVID-19.

1.2. Banco de Guatemala (BANGUAT)

Entre los años 1924-1926 en Guatemala ocurrió la reforma monetaria y financiera, con dicho acontecimiento se tuvo la creación del Banco Central de Guatemala como establecimiento de emisión, giro y descuento de carácter privado y con participación del Estado como accionista. Esta reforma se vincula a grandes desequilibrios monetarios y financieros por gobiernos anteriores. En noviembre de 1924 fue promulgada la Ley Monetaria de la Republica de Guatemala que daría vida a la nueva unidad monetaria: Quetzal.

Con la gran depresión mundial de los años 1929-1933 se afectó gravemente la economía de Guatemala, lo que causo conflictos en el Banco

4

Central y su política monetaria, por lo que esto conllevo a impulsar una nueva reforma monetaria y bancaria entre 1944-1964, mediante la cual finalmente se creo el Banguat como sucesor al antiguo Banco Central de Guatemala. Cabe destacar que la revolución de 1944 aporto con otorgarle al Banguat la calidad de banco estatal y facultad realizar una policita monetaria.

La estructura básica de la legislación financiera guatemalteca de 1945 y 1946 permitió el funcionamiento ordenado durante los primeros 40 años, sin embargo, a finales de los años ochenta el impacto de crisis económicas regionales, la liberación de la banca y de os mercados financieros internacionales, el avance en la informática, electrónica y comunicaciones rebasaron las ideas que fueron tomadas en los años cuarenta. Por lo que surge la necesidad de modernizar el sistema bancario y en el año de 1993 surge el Programa de Modernización del Sistema Financiero Nacional.

El programa de modernización incluyo una serie de medias que adoptaron la Junta Monetaria y el Congreso de la Republica y los Ministerios de Estado. Fueron mas de 50 resoluciones en donde se vieron afectadas políticas crediticias, liberalización y diversificación de productos y servicios bancarios y funcionamiento de supervisión financiera. Entre las disposiciones mas destacadas se tienen:

- Acuerdo Legislativo No. 18-93: Prohibición de que el banco central financie al gobierno.
- Decreto 12-95: Reforma a Ley Orgánica del Banguat para fortalecer la capacidad supervisora de la Superintendencia de Bancos.
- Decreto 23-95: Reforma Ley de Bancos
- Decreto 29-95: Liberación de contratación de tasas de interés.

 Decreto 44-95: Reforma la Ley de Bancos de Ahorro y Préstamo para la Vivienda Familiar.

Finalmente, para consolidar los cambios de los años noventa el 1 de junio del año 2000 se formalizaron los lineamientos de una reforma integral que pretendía un cambio en la concepción del papel de la banca central y en la orientación de la regulación financiera, por medio de una serie de medias clasificadas en tres grandes áreas de acción:

- Bases para la reforma integral: Elaboración de un diagnóstico de la situación del sector financiero que permita preparar las bases para las reformas legales estructurales.
- Reforma integral a las leyes financieras
- Modificaciones reglamentarias para desarrollar el contenido de la reforma integral.

El resultado de dicha reforma integral se centró en la elaboración de cuatro proyectos fundamentales:

- Ley Orgánica del Banguat: como fin principal fortalecer la estabilidad macroeconómica que permita a los agentes económicos la correcta toma de decisiones referentes al consumo, ahorro e inversión productiva.
- Ley Monetaria: responsabilidades de la emisión monetaria.
- Ley de Bancos y Grupos financieros: Orientada a propiciar la estabilidad del sistema bancario para fortalecer los pagos y aumentar la solidez y solvencia del sistema financiero.
- Ley de Supervisión Financiera: orientada a favorecer la solidez y solvencia del sistema bancario y promover la confianza del publico en el sistema bancario.

Dichas leyes fundamentales obtuvieron su vigencia el 1 de junio de 2002. Con este contexto histórico obtenido de Banguat⁴, vemos el por qué actualmente Guatemala posee en su economía una de las mejores regulaciones de Latinoamérica, y por que es que el Quetzal no se ha visto devaluado de manera exorbitante a comparación de otras monedas, y es que gracias a estas reformas que le brindaron la autonomía al Banguat, pueda regular el ámbito económico del país, y que este genere una inflación que ponga en peligro el estatus que tiene su moneda.

1.3. Actividades económicas de Guatemala

El PIB de Guatemala está dividido principalmente en 17 sectores económicos⁵, estos sectores son los que representan como está principalmente distribuida la economía en el país. Con el pasar del tiempo, han ido cambiando la forma en que están distribuidas las actividades, ya que la economía se encuentra en constante evolución y surgen nuevas necesidades en cada país.

Guatemala ha tenido cambios en respecto a la relevancia de sus actividades económicas, y dependiendo de la situación en la que se encuentre el país, se observa un crecimiento o disminución en la actividad. Entre las principales actividades económicas tomadas del Top 5 que serán tomadas para el análisis en el presente trabajo de investigación se tienen:

⁴ Banguat. *Reseña Histórica*. https://www.banguat.gob.gt/es/page/resena-historica. Consultado: octubre 2021.

⁵ Banguat. *Estudio de la Economiía Nacional 2020.* https://www.banguat.gob.gt/sites/default/files/banguat/Publica/doctos/estudio_de_la_economia_2020.pdf. Consultado: octubre 2021.

- Agricultura, ganadería, silvicultura y pesca: sector que se conforma principalmente por la producción agrícola, destacan entre ellas la producción y exportación de banano, cardamomo, café y caña de azúcar.
- Comercio y reparación de vehículos: conformado por la producción de bienes agropecuarios e industriales, importaciones de bienes en términos reales como: la gasolina, gas, Diesel, gas propano, maíz, bebidas, líquidos alcoholismo y vinagres entre otros.
- Industrias manufactureras: conformado por la demanda interna y externa que surgen de materiales de construcción, productos de plástico, productos químicos y productos de papel. Asimismo, lo conforma el grupo de la fabricación de textiles y predas de vestir, elaboración de aceites y grasas de origen vegetal y animal.
- Actividades Inmobiliarias: actividad conformada por la cantidad de viviendas terminadas que pasan a formar parte del inventario disponible de viviendas del país y de los locales comerciales que son demandados por el resto de las actividades económicas.
- Información y comunicaciones: el valor agregado de dicha actividad es por las telecomunicaciones, que va en constante crecimiento por la demanda de contratación en telefonía móvil, expansión del comercio y la publicidad digital y la necesidad de mayor acceso de banda ancha en los servicios de internet.

1.4. Producto Interno Bruto

El Producto Interno Bruto -PIB- es el valor monetario de los bienes y servicios finales producidos por un territorio durante un periodo determinando, por lo general es en un año. El propósito que tiene es de medir el nivel de riqueza de un país en un lapso definido, por lo que nos puede decir si esta esta en crecimiento o si ocurre el efecto contrario que es una caída.

8

1.4.1. Cálculo de PIB

El PIB se calcula en base a tres métodos o enfoques diferentes: según la producción, según el flujo de datos o según el flujo de rentas. Cualquiera de estos metidos debe coincidir con el resultado final, y el uso de estos métodos dependerá de los datos que se tengan disponibles.

1.4.1.1. Según la producción o el valor añadido

Es un cálculo donde totalizamos el valor de las ventas de los productos conocido como Valor añadido bruto VAB menos el valor de las materias primas, más la diferencia entre impuestos y subvenciones. Teniendo como resultado la formula:

Este calculo pretende evitar la doble contabilización de bienes, ya que algunos de estos pueden ser un recurso, pero a la vez un intermediario para un producto final, lo que causaría que se estuviera tomando de manera duplicada.

1.4.1.2. Según el flujo de gastos

Se calcula en base al valor de las adquisiciones o consumos realizados por usuarios finales, ya que gracias a este flujo se puede conocer y evaluar las razones en las que se puede tener una caída cuando hay una disminución del consumo interno de los habitantes de un país. El cálculo se realiza de la forma siguiente: PIB = Remuneración de los trabajadores + Excedente bruto de explotación + (impuestos – subvenciones)

1.4.1.3. Según los ingresos o flujo de rentas

El punto central de este calculo es el total de ingresos que resultad de la actividad productiva, y este visibiliza la repartición de las rentas entre los agentes productivos. Cuando la producción es vendida, se generan ingresos en las empresas y estos se distribuyen a través de los salarios, que a su vez representan ingresos para las familias. La formula para el cálculo es:

PIB = CONSUMO + INVERSIÓN + GASTO PUBLICO + EXPORTACIONES + IMPORTACIONES

2. CIENCIA DE DATOS

2.1. Ciencia de datos

"La ciencia de datos es un enfoque multidisciplinario para extraer conocimientos prácticos de los grandes y cada vez mayores volúmenes de datos recopilados y creados por las organizaciones de hoy. La ciencia de datos abarca la preparación de datos para su análisis y procesamiento, la realización de análisis de datos avanzados y la presentación de los resultados para revelar patrones y permitir que las partes interesadas saquen conclusiones informadas."⁶

Para la preparación de estos se implica que se realicen procesos de limpieza, agregación y manipulación de los datos para que estén listos para su procesamiento. Para el análisis son requeridos algoritmos, análisis y modelos de inteligencia artificial. Y la forma de presentar los resultados debe ser por medio de herramientas de visualización de datos que permitan ver los patrones y comprender las tendencias de las predicciones realizadas.

2.1.1. Ciclo de vida de la Ciencia de Datos

También llamado canalización de la ciencia de datos incluye ciertos procesos continuos en los que podemos incluir:

• Captura: es la recopilación de datos estructurados y no estructurados sin procesar de todas las diversas fuentes de donde provenga la información.

⁶ International Business Machines Corporation. *Ciencia de los datos.* https://www.ibm.com/cloud/learn/data-science-introduction. Consulta: septiembre de 2021.

- Preparar y mantener: implicar poner los datos sin procesar en un formato coherente para análisis. En este proceso puede ser por medio de ETL u otros procesos de integración de datos y tenerlos en un almacén de datos para posterior tener el análisis.
- Proceso: se examinan los sesgos, patrones, rangos y distribuciones de valores dentro de los datos para que estos puedan ser utilizados para su respectivo análisis predictivo.
- Analizar: en esta parte es donde se obtiene un descubrimiento, y donde el científico de dato lo realiza por medio de análisis estadísticos, análisis predictivos, regresiones o por medio de otros métodos.
- Comunicar: los resultados se presentan por medio de reportes, gráficos u otras maneras de visualización y el impacto que pueden causar estos resultados.

2.2. Analítica de Datos

Por medio de la analítica de datos se pueden obtener los datos para ser procesados, por lo que entender desde su definición y diversos tipos enfatizará mejor en la importancia de ésta.

2.2.1. Definición de análisis de datos

Es la ciencia de analizar datos sin procesar para sacar conclusiones sobre esa información por medio de diversos algoritmos y procesos. Por medio de los análisis se pueden optimizar procesos, mejorar rendimientos e identificar formas mas eficientes de operar. Entre los pasos para realizar un análisis de datos se tienen:

- Determinar los requisitos de datos o la forma en que se agrupan los datos.
- Recopilar la información por medio de diversas fuentes como computadoras, entrevistas, fuentes en línea o personales.
- Organizar la información para que pueda ser analizada.
- Limpiar los datos para asegurarse que no hayan duplicados o errores, y así poder ahorrar cualquier futura faya en el análisis.

2.2.1.1. Tipos de análisis de datos

Existen diversos tipos de análisis de datos, dependiendo de las fuentes de información o del tipo de pronóstico que se desee obtener, por lo que entender cuáles son, mejorará el tipo de análisis a realizar. Por lo cual, se presentan los principales tipos de análisis que existen.

2.2.1.1.1. Análisis descriptivo

Es el análisis que describe lo sucedido durante un periodo de tiempo determinado y así responder las preguntas sobre lo ocurrido. Este tipo de técnica resume grandes conjuntos de datos para describir los resultados a las partes interesadas.

2.2.1.1.2. Análisis de diagnóstico

Se enfoca más en encontrar la razón de algo que sucedió, por lo que se tienen entradas de datos más diversas y un poco de hipótesis. Con estas técnicas se complementa el análisis descriptivo más básico, ya que toman los hallazgos del descriptivo y profundizan en encontrar la causa. Este tipo de análisis ocurre generalmente en tres pasos:

13

- Identificar las anomalías de los datos
- Recopilar datos relacionados con las anomalías.
- Utilizar técnicas estadísticas para encontrar relaciones y tendencias que puedan explicar las anomalías.

2.2.1.1.3. Análisis predictivo

Se traslada a lo que probablemente sucederá a corto plazo. Estas técnicas utilizan datos históricos para identificar tendencias y determinar si es probable que se repitan. Las herramientas predictivas ayudan a brindar información valiosa sobre lo que puede suceder en un futuro por medio de técnicas estadísticas y de aprendizaje automático.

2.2.1.1.4. Análisis prescriptivo

Este tipo de análisis sirve para poder responder preguntas acerca de lo que se debe de hacer. Con su uso se pueden tomar decisiones basadas en datos. Las técnicas utilizadas se basan en estrategias de aprendizaje automático que pueden encontrar patrones en grandes conjuntos de datos.

2.2.2. Integración de datos

La integración de datos es la práctica de consolidar datos de fuentes dispersas en un solo conjunto de datos con el objetivo de brindar la información de forma consolidada y exacta que satisfaga un fin. Esta práctica es uno de los componentes principales del proceso general de gestión de datos, y cada vez es mayor su uso en big data. Entre las principales técnicas de integración de datos se tienen:

- Extraer, transformar y cargar: los datos se recopilan, armonizan y carga en un almacén de datos o base de datos.
- Extraer, cargar y transformar: los datos se cargan tal cual en un sistema de big data y se transforman posteriormente para los usos analíticos.
- Captura de datos de cambios: identifica los cambios de datos en las bases de datos en tiempo real y los aplica en un almacén de datos.
- Replicación de datos: los datos en una base de datos se replican en otras bases de datos para mantener la información sincronizada con los otros usos operativos y para la copia de seguridad.
- Virtualización de los datos: los datos de diferentes sistemas se combinan virtualmente para crear una vista unificada en lugar de cargar datos en un nuevo repositorio.
- Streaming Data Integration: un método realizado en tiempo real en el que diferentes flujos de datos se integran continuamente y alimentan a los sistemas de análisis y almacenes de datos.

2.2.2.1. Herramientas y técnicas de Integración de datos

Estas herramientas están disponibles en una amplia gama de niveles organizativos, entre los más típicos se tienen:

- Integración manual: no hay vista unificada de los datos. El usuario es el encargado de operar toda la información relevante por medio de acceso a los sistemas.
- Integración basada en aplicaciones: requiere que una aplicación implemente todos los esfuerzos de integración.
- Integración de datos de middleware: transfiere la lógica de una aplicación a una nueva capa de middleware.

- Acceso uniforme a los datos: deja los datos en sistemas de origen y define un conjunto de vistas para proporcionar una vista unificada.
- Almacenamiento de datos común: crea un nuevo sistema en el que se almacena y gestiona una copia de los datos del sistema de origen independiente del sistema original.

2.3. Lenguaje R

R es un lenguaje y ambiente hecho para la computación y gráficos estadísticos. Provee una amplia variedad de técnicas estadísticas, pruebas estadísticas, análisis de series de tiempo y técnicas graficas. R posee una gran facilidad con la que se pueden producir gráficos con calidad de publicación bien diseñados, incluidos símbolos matemáticos y formulas.

2.3.1. Entorno R

R es un conjunto integrado de instalaciones en el software para poder manipular, calcular y graficar datos, que incluye las siguientes características:

- Un efectivo manejo de información y fácil almacenamiento.
- Un conjunto de operadores para cálculos en matrices.
- Una amplia colección de herramientas intermedias para en análisis de datos.
- Facilidades graficas para el análisis de datos y que puedan ser visualizadas en pantalla.
- Un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas e instalaciones de entrada y salida.

2.3.2. RStudio

RStudio es un Entorno de Desarrollo Integrado -IDE- para R. Incluye una consola, un editar de resaltado de sintaxis que admite la ejecución directa de código, así como herramientas para trazar, historial, depuración y gestión del espacio de trabajo.

5			
🚯 Ritudio			- 0 X
File Edit Code View Plots Session Build Debug Profile Tools Help		,	
0 • 00 @ • 0 00 0 + 6 to to feature 0 • Addis •			Project (Nore) •
Console Terminal = Jobs =	Environment History Connections Tutorial		-7
@ 8410 / @	/ 🐨 🔒 🐨 Import Dataset + 🐊 79 MB + 🖌		= ue • 😳 •
R version 4.1.0 (2021-05-18) "Camp Pentanezen" Copyright (C) 2021 The R Foundation for Statistical Computing Platform: X&de Heiden Height 2026 (de-hit)	A + 🖓 data Inviorment +		۹.
R is free software and comes with ABSOLUTELY ND WARRANTY. You are welcome to redistribute it under certain conditions. Type "license") or "licence") for distribution details.	Environment is empty		
<pre>R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citatien()' on how to cite k or # a packages in publications.</pre>			
Type (demo() for some demos, 'help()' for on-line help, or 'help,start()' for an HTML browser interface to help. Type 'd()' to quit R.			
>1			
	Nes Plots Packages Help Viewer		
	😟 New Folder 👂 Delete 🗼 Renarre 🔅 Mare +		
	C & Home		
	A Nane	Size	Modified
		10.5.49	Are 20 2020 1162 444
		10.2 1.0	Apr an, anno, i ture Arm
	201801074, Yates accelb	2.1 MB	May 2, 2020, 11:13 PM
	Android Projects		
	🗆 🧉 Apowenati		
	📄 📫 Apoyo Compi 2		
	🗌 🧰 Archivos de Prueba		
	 archivos_calificacion_proyecto1 		
	🗌 🤷 Arduino		
	ArduineData		
	 Autodesk Application Manager 		
	Aure		
	D D Make MT-miletable	207.0	1.120 2010 0.00 201
	Beckup CP Teres T.pdsbek	8 C65	Jul 30, 2018, 840 PM
	Banfort		
	Reciment Petion		
	Calculadora		
	Carta Sentiago do x	99.9 XB	Jun 14, 2020, 11:37 PM
	🗇 🧰 Clases de Progra		
	Corriandos de Git.docx	180.5 KB	Feb 13. 2020. 1:40 PM *

Figura 1. RStudio

Fuente: elaboración propia, empleando RStudio.

2.4. Análisis de tendencias a nivel de analítica de datos

Los siguientes métodos realizan pronósticos de tendencias en base a la analítica que se realiza de los diversos datos. La diferencia entre ellos es la forma en que realizan el cálculo de la tendencia.

2.4.1. Método Naive

Es el método más simple, se basa principalmente en: el pronóstico para mañana es lo que estamos observando hoy. Puede ser implementado en funciones homónimas que llega a ser útil para datos de precios de acciones, e inclusive si no es el método ideal proporciona un punto de referencia útil para métodos de pronósticos. Tienen la ventaja que es económico de desarrollar, almacenar datos y operar; pero con la desventaja de que no considera posibles relaciones causales. Estos modelos se pueden clasificar en dos grupos:

- Modelos de proyección simple: requieren entradas de datos de observaciones recientes, pero no se realiza ningún análisis estadístico.
- Modelos complejos: estos contienen la suficiente complejidad que llegan a requerir una computadora.

2.4.2. Método Exponential Smoothing

Es un método de pronostico de series de tiempo para datos univariados donde el pronostico se realiza a través del promedio ponderado exponencialmente de observaciones anteriores. Tiene el objetivo de suavizar las series originales de la misma manera que lo hace el promedio móvil y hacer uso de datos suavizados para pronosticar los valores futuros de la variable de interés. Entre los diferentes tipos se tienen:

- Simple Exponential Smoothing: sirve para series de tiempo para datos univariados sin una tendencia o estacionalidad.
- Double Exponential Smoothing: es una extensión del simple que agrega explícitamente soporte para las tendencias en la serie de tiempo univariante.

 Triple Exponential Smoothing: es una extensión que agrega explícitamente soporte para la estacionalidad a la serie de tiempo univariante.

2.4.3. BATS Y TBATS

Ambos son métodos de pronóstico de series de tiempo, que son capaces de modelar series de tiempo con múltiples estacionalidades y ser implementados en la clase principal de estimador.

2.4.3.1. TBATS

Es un modelo basado en el Exponential Smoothing en donde cada estacionalidad esta modelada por una representación trigonométrica basada en series de Fourier. Tiene la ventaja de que solamente requiere dos estados independientemente de la duración del periodo y la capacidad de modelar los efectos estacionales de longitudes no enteras.

2.4.3.2. BATS

La diferencia que posee con TBATS es la forma en la que modelan los efectos estacionales, con un enfoque mas tradicional por lo que solamente puede modelar longitudes de periodos enteros.

2.4.4. Modelos ARIMA / SARIMA

ARIMA y SARIMA son algoritmos de pronósticos, ambos toman en cuenta valores del pasado y realizan pronóstico del futuro, con la diferencia de que SARIMA toma en cuenta cualquier patrón de estacionalidad.

2.4.4.1. ARIMA

Es un modelo que explica una serie de tiempo determinada en función de sus propios valores pasados, es decir, sus propios retrasos y los errores de pronóstico retrasados con lo que puede utilizar la ecuación para pronosticar valores futuros.

2.4.4.2. SARIMA

Es una extensión de ARIMA que admite explícitamente datos de series de tiempo univariantes con un componente estacional. Agrega tres hiperparámetros para especificar la auto regresión -AR-, la diferenciación -I- y la media móvil -MA-para el componente estacional de la serie, así como un parámetro adicional para el periodo de la estacionalidad.

3. ANALÍTICA Y VISUALIZACION DE DATOS

3.1. Microsoft Power Bi

"Microsoft Power Bi es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente."⁷

Microsoft Power Bi consta de varios elementos que funcionan de manera conjunta, y se resume en tres conceptos básicos:

- Aplicación de escritorio de Windows Ilamada Microsoft Power Bi Desktop.
- Servio SaaS (software como servicio) en línea denomado servicio Microsoft Power Bi.
- Aplicaciones móviles de Microsoft Power Bi para dispositivos.

⁷ Microsoft. ¿Qué es Microsoft Power Bi? https://docs.microsoft.com/es-es/powerbi/fundamentals/power-bi-overview. Consulta: 1 de octubre de 2021.



Fuente: Microsoft. ¿Qué es Microsoft Power Bi? thttps://docs.microsoft.com/es-es/powerbi/fundamentals/media/power-bi-overview/power-bi-overview-blocks.png. Consulta: 1 de octubre de 2021.

3.1.1. Tecnologías de Microsoft Power Bi

Para realizar un mejor análisis en Microsoft Power Bi, existen diversas herramientas que realizan funciones específicas que procesan la información para óptimos resultados.

3.1.1.1. Power Query

Es un motor de transformación y preparación de datos. Viene con una interfaz gráfica para obtener datos de fuentes y un editor para aplicar transformaciones. Con esta tecnología se puede realizar el procesamiento de datos de extracción, transformación y carga. Existen cuatro fases para usar Power Query:

 Conectar: realiza conexiones a datos en la nube, en un servicio o de forma local.

22

- Transformar: da forma a los datos para satisfacer sus necesidades, mientras que la fuente de información permanezca sin cambios.
- Combinar: integra datos de varios orígenes para obtener una vista única en los datos.
- Cargar: completa la consulta y la carga en una hoja de cálculo o en un modelo de datos.



Figura 3. Fases de Power Query

Fuente: Microsoft. *Acerca de Power Query en Microsoft Excel.* https://support.content.office.net/es-es/media/e5c63e6e-3205-4803-8b4a-6e469a211089.png. Consulta: 2 de octubre de 2021.

3.1.1.2. DAX

Data Analysis Expressions -DAX- es un lenguaje de expresión de fórmulas que se usa en Microsoft Power Bi. Las fórmulas incluyen funciones, operadores y valores para realizar cálculos y consultas avanzados sobre datos en tablas y columnas relacionados a modelos de datos tabulares y son utilizadas en:

 Medidas: son fórmulas de cálculo dinámicas donde los resultados cambian según el contexto. Se utilizan en informes que admiten la combinación y el filtrado de datos mediante el uso de varios atributos como un informe de Microsoft Power Bi.

- Columnas Calculadas: es una columna que agrega a una tabla existente y luego crea una formula DAX que define los valores de columna.
- Tablas Calculadas: es un objeto calculado, basado en una expresión de formula, que, en vez de consultar y cargar valores en las columnas desde una fuente de datos, se realiza por medio de una formula DAX.
- Seguridad a nivel de fila: una fórmula de DAX debe evaluar una condición de VERDADERO/FALSO, definiendo que filas pueden ser devueltas por los resultados de una consulta.
- Formulas: es la parte esencial para crear cálculos en columnas y medidas calculadas, y proteger datos mediante la seguridad a nivel de fila.

4. HERRAMIENTA PARA EL ANÁLISIS DE PRONÓSTICOS

4.1. ETL de datos

Extraer, transformar y cargar -ETL- es una parte de la integración de datos que consiste en 3 principales fases las cuales están descritas en sus siglas. El fin de este proceso es que al realizarlo los datos ya se encuentren limpios y listos para ser utilizados evitando que se tenga información corrupta que pueda afectar el resultado del análisis.

4.1.1. Extracción de datos

Para el proceso de extracción se mostrará paso a paso cómo procesar la información desde la fuente de origen hasta tenerlo en un dataset.

4.1.1.1. Descarga de la fuente de datos de Banguat

La obtención de datos es por medio de la pagina centra del Banguat. Los pasos para poder tener acceso a la información son los siguientes:

- Ingresar a la pagina de Banguat (https://banguat.gob.gt/es)
- Ingresar al apartado de Estadísticas Macroeconómicas.
- Seleccionar Actividad Económica.
- Posterior, seleccionar Producto Bruto Interno Trimestral.
- Luego seleccionamos Sistema de Cuentas Nacionales Trimestrales Año de Referencia 2013.

 Finalmente seleccionamos Graficas y Cuadros Estadísticos para ver en Formato Exel.

Finalmente tendremos el siguiente cuadro con la información referente al PIB de Guatemala.

1	iutoguarda	to 💽 💽	୭୦୧୦୫	~ ~ C	NT_cuadros_al_2T	_2021 <i>x</i> lsb - Exc	el 🔎	Buscar				a de la composición de la comp		Herbert	Reyes HR	• - • •	o ×
Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Programador Ayuda Team 🖄 Compartir 🖓 C											artir 📃 🖓 Com	nentarios					
Pe	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $											denar y Buscar iltrar ~ seleccion	y Analizar ar ~ datos				
	$1 \rightarrow 1 \times \sqrt{k}$ (and 1)																
Al		• • · · ·	√ Jr Cu	Jadro 1													Ŷ
1	A	В	С	D	E	F	G	H	I) 	K adro 1	L	M	N	0	Р	Q ^
2		PRODUCTO INTERNO BRUTO TRIMESTRAL (PIBT), MEDIDO POR EL ORIGEN DE LA PRODUCCIÓN											_				
3										PERÍODO: 1	T-2013 - 2T-2	021					
4										Millones	de quetzales						
5		A	в	c	D-E	F	G	н		J	к	L	м	N	0	Р	0
7	Período	Agricultura, ganadería, silvicultura y pesca	Explotación de minas y canteras	Industrias manufac- tureras	Suministro de electricidad, agua y saneamiento	Construcción	Comercio y reparación de vehículos	Transporte y almacena- miento	Actividades de alojamiento y de servicio de comidas	Información y comunica- ciones	Actividades financieras y de seguros	Actividades inmobiliarias	Actividades profesionales científicas y técnicas	Actividades de servicios administra- tivos y de apoyo	Administra- ción pública y defensa	Enseñanza	Saluc
8	2013	43,121.6	5,826.7	60,027.6	11,882.6	17,835.4	76,016.4	11,073.5	11,492.3	19,122.5	14,001.8	36,214.0	10,343.1	14,549.2	16,754.2	17,803.1	9,1
9	1	12,146.3	1,548.3	15,323.7	2,779.6	4,241.6	18,902.5	2,796.1	2,622.1	4,433.7	3,509.1	8,867.0	2,419.6	3,653.1	3,874.1	4,744.2	2,1
10	Ш	10,728.4	1,380.0	15,018.0	2,754.5	4,532.2	18,689.7	2,747.1	2,809.3	4,764.9	3,397.1	9,006.6	2,566.9	3,539.4	3,717.3	4,085.5	2,1
11		9,919.6	1,338.5	14,226.4	3,051.7	4,383.2	18,567.8	2,643.9	2,899.1	4,687.9	3,477.0	9,090.5	2,528.7	3,354.9	4,362.1	4,880.8	2,:
12	IV	10,327.2	1,559.8	15,459.5	3,296.8	4,678.4	19,856.5	2,886.3	3,161.8	5,235.9	3,618.7	9,250.0	2,828.0	4,001.9	4,800.8	4,092.7	2,!
13	2014	45,065.8	7,372.7	62,968.6	12,541.3	20,864.6	82,963.0	12,090.5	12,390.8	19,434.9	15,520.9	38,053.4	10,913.5	15,547.8	17,930.8	19,559.0	10,1
14	1	11,718.8	1,720.9	15,909.2	3,073.3	4,998.1	19,980.1	2,815.1	2,793.7	4,643.0	3,956.5	9,352.1	2,608.1	3,833.7	3,964.8	5,170.8	2,1
15		10,931.0	1,916.4	15,654.3	3,070.9	5,299.7	20,021.5	2,859.2	3,072.1	4,841.5	3,754.6	9,478.2	2,663.8	3,852.5	4,046.5	4,516.0	2,4
16		10,928.3	2,089.7	14,894.1	2,938.3	5,161.8	21,108.3	2,950.9	3,062.7	4,761.0	3,836.1	9,523.6	2,645.5	3,687.7	4,754.2	5,367.9	2,1
17	IV	11,487.8	1,645.7	16,511.0	3,458.8	5,404.9	21,853.1	3,465.4	3,462.3	5,189.3	3,973.7	9,699.4	2,996.1	4,173.9	5,165.3	4,504.3	2,1
18	2015	47,445.2	5,571.7	68,196.6	11,281.9	23,185.6	86,989.8	14,565.7	13,867.5	20,814.2	16,598.9	40,138.3	11,612.9	16,703.5	19,184.2	21,770.9	11,2
19	1	13,068.1	1,609.8	17,109.6	3,147.6	5,123.1	21,043.5	3,595.2	3,155.9	4,859.4	4,147.1	9,852.1	2,871.1	4,269.7	4,274.1	5,834.5	2,1 🗸
		1 2 3 4	5 6 7	8 +							-						Þ

Figura 4. Microsoft Excel información PIB

Fuente: elaboración propia, con datos obtenidos de Banguat 2021.

4.1.1.2. Vinculación de la data con Microsoft Power Bi

Para poder realizar la vinculación entre Microsoft Power Bi con la data que tenemos generada en Microsoft Excel, necesitamos primero iniciar la aplicación. Una vez iniciada veremos que tendremos directamente la opción de cargar información desde Microsoft Excel.

Figura 5. Inicio de Microsoft Power Bi

ଅଚେଡ		Untitled - Po	ower BI Deskt	op			Q	Search													Sign in 🔵 🛛 🗕	6	×
File He	lome Insert	Modeling	View	Help																			
Paste Cut Que Cop	py G mat painter dat	t Excel Po	ower BI SQL atasets Serve	Enter Dat data	taverse Re sou	ecent rces v	Transform Rel data ~	sfresh N	Vew To	A di iext Mi xxx visu	ore als ~	New measure	Quick measure	Sensitiv	ity Pi	ublish							^
bel.																				<	Visualizations	>	<
=																				\forall			Ŧ
60	Add data to your report											Filters			lds								
														R = R -4 □ 2 Ω	Py E ₩⊗								
		Once loaded, your data will appear in the Fields pane.											Values 7										
				X			10			C	ò			Q							Add data fields her	e	
																					Drill through		
			le .	nport data fror	m Excel	leg	port data from SC	QL Server	Paste	e data into	a blank to	table	Try a samp	ile dataset							Cross-report		
							Get d	data from	anothe	r source	·										Keen all filters		
																					On —●		
																					Add drill-through fi	elds here	
< >	Page 1 +																						
Page 1 of 1	_																						

Fuente: elaboración propia, empleando Microsoft Power Bi.

Seleccionamos la opción para poder importar desde Microsoft Excel, luego nos pedirá que ubiquemos donde esta nuestro archivo de Microsoft Excel para posteriormente seleccionarlo. Luego de realizar el proceso, nos pedirá que seleccionemos la hoja de Microsoft Excel en donde tenemos la información.

Figura 6. Carga de Microsoft Excel en Microsoft Power Bi

File Home Insert Modeling View		P. D. Barnester						54	pin 🔾 — 🗗	
	Help						254			
	Navigator					D X				
Clipboard		Q	Hoja1			G.				-
A There are pending changes in your queries that	Display Options =	Da	ID Alle	Periodo	Categoria		ed changes	×		
TT notes a point of any set of the set of the	a 📕 Dates Taxis site (11		1	2025 Total Periodo	Agricultura, ganaderia, silvicultura y pesca		re charges	~ 1	/isualizations	
1			2	2025 1	Agricultura, ganaderia, silvicultura y pesca	^		. 8		-
	Co- Hoja i		3	2025 #	Agricultura, ganaderia, silvicultura y pesca			1	× (A) he has his het	10
5			-4	2025 11	Agricultura, ganaderia, silvicultura y pesca			1		g.
			.5	2035 N	Agricultura, ganaderia, silvicultura y pesca			ers E		
			6	2025 Total Periodo	Comercio y reparacion de vehiculos					
			7	2015 1	Comercio y reparacion de vehiculos			1		
			. 8	2025 10	Comercio y reparacion de vehiculos			-		
			9	2025 11	Comercio y reparacion de vehiculos			2	···· (
			10	2025 N	Comercio y reparacion de vehiculos				m 9	
			11	2025 Total Periodo	industrias manufactureras				CCC 0	
			3.2	2015 1	Industrias manufactureras			1	Talues .	
			13	2025 #	Industrias manufactureras				Add data fields here	
			14	2025 11	industries manufactureras					
			15	2025 N	industrias manufactureras			21	Drill through	
			31	2015 Total Periodo	Actividades inmobiliarias			1		
			17	2015 1	Actividades inmobiliarias			· ·	ross-report	
			18	2015 #	Actividades inmobiliarias				Off O-	
			2.9	2025 11	Actividades inmobiliarias					
			20	2025 N	Actividades inmobiliarias				eep all litters	
			21	2025 Total Periodo	Informacion y comunicaciones				On	
			22	2025 1	informacion y comunicaciones					
			23	2025 1	Informacion y comunicaciones	~		12	Add drill-through fields here	
			<		3	>				

Fuente: elaboración propia, empleando Microsoft Power Bi.

Finalmente seleccionamos la opción de cargar y tendremos la información cargada, lista para poder ser manipulada en Microsoft Power Bi.

4.1.2. Transformación de datos

En la fase de transformación, se realizarán diversos procesos para tener como resultado información con calidad que pueda brindar mejores pronósticos.

4.1.2.1. Limpieza, filtrado, transformación e inclusión de la fuente de datos

Para la transformación de los datos se realizo una transformación de la información en una forma lineal, con los siguientes campos principales:

- ID: dato numérico único de cada conjunto de registros que se tendrán.
- Año: el año del que se esta tomando el dato.

- Periodo: es el periodo trimestral o el resultado de la suma de los 4 periodos trimestrales del año en que ocurre.
- Categoria: se refiere a la actividad económica.
- Valor Q(M): es el valor en millones de quetzales de la categoría en el periodo y año de tiempo definido.

4.1.2.1.1. Transformación automática por medio de Power Query

Cuando se desea realizar la transformación automática es que se realiza un cambio en general para toda la data, de manera que no sea necesario realizar el cambio dato por dato. Para realizarlo seleciconamos en el menú de Herramientas de Tabla, la opción de Transformar Información.

Figura 7. Transformar con Power Query

6	C Utified Rever B Desttop Destroy		Sign	ы ○ – ах
File	Home Insert Modeling View Help Table tools			
Easte				
	of formit printin data - workbook datatets Server data sources data - wixed box violatis - metsour neoure neoure recurs - Clobard Data - TTT - rent Clobarions Serving Sure			^
ial.	De manotem data Costa source estings	<	Visualizations >	Fields >
	Eißt parameters	4		P Search
€8	Edit variables	Filter		✓ I General
		, v	© ₩ ∞ □ E ⊡ □ □ ■ R Py E	Categoria
				□ ∑ ID □ Periodo
			87	□ ∑ Valor Q(M) >
	Duild double with some date		Volues Add data fields have	
	Build Visuals with your data		Drill through	
	Select or drag fields from the Fields pane onto the report canvas.		Cross-report	
			011 O	
			Keep all filters	
			Add drill-through fields here	
	Page 1 4			
Page 1 of	e1			

Fuente: elaboración propia, empleando Microsoft Power Bi.

En esta seccion podremos ver los datos, si se encuentran con error, vacíos o la validez que tienen cada uno de ellos. Por medio de Power Query podremos realizar la limpieza de los datos, por si hubiera datos que puedan afectar la calidad de la información.

Manager Transformer					
Interest Interested	Add Column View Tools Help				
🗋 🗋 🖽	1 🕼 🔢 🔓 Prope	atan 🔟 🖌 🎬 🖬 🖞 🚠 🖻	Data Type: Whole Namber * In Marge Queries * In Use Fest Row as Headers * T Append Queries *	The tradytes Original Distance	
8 New Recent Enter	r Data source Manage Refresh Manag	of Choose Renova Keep Renova Split Grou	P 1 Replace Values	A Azure Machine Learning	
Nee Garry	Data Sources Parameters Query	Manage Columns Reduce Rows Sort	Transform Combine	Al models	
	This mendant may be up to 2 days out	Referit			A man france
iona (r)	·				dreak serruda
General	🔀 🗸 🏂 = Table.TransformCole	umTypes(#"Promoted Headers",{{"ID", Int64.Type}, {"AMo",	, Int64.Type}, {"Periodo", type text}, {"Cat-	regoria", type text), ("Valor Q(R)", type number}})	V + PROPERTIES
Top 5	- H. M.	- A factoria	- 1.2 Volus 01Mi	5	Name
	a block the second second		VAND - Valor Anno	-	General
	Error D% Error	D% Error D% Error	0% Error 0%		All Properties
	Empty D% Empty	0% • Empty 0% • Empty	D% Empty D%	1	A ADDRESS OF STREET
		100 0000000		(· APPLIED STEPS
		(Source
		(1001 00000000000000000000000000000000		1	Provident Maladam
	776 Bratines, 776 unique V distinct, 6 u	Index 2 destruct 2 under 18 district 0 under	774 distinct, 774 unique		> Channel Type
	1	202 Y T Seriedburg on	aderia, silviout. 17146.1944		^
	3	2023 12 Aaricultura, can	aderia, allwicult		
	4	2013 III Burimaltura, can	aderia, silvicala	in the second	
	5 7	2023 IV Agriculture, gan	aderia, milvicult. 20327.2166	ia l	
	6 /	2014 Total Periodo Agricultura, gan	aderia, silvicult_ 45065.8169		
	7 7	2014 I Agricultura, gan	aderia, allvicult. 11718.7909	и	
	A	J014 II Agricultura, gan	inderia, milvioult 10930.5610		
	9 2	2014 III Agricultura, gan	aderia, silvicult. 10928.2791	a .	
	10 2.0	2014 TV Agriculture, gen	inderia, allvioult		
	11 22	2015 Total Periodo Agricultura, gan	aderia, allvicult. 47443.2096		
	12 22	2015 1 Agriculture, gen	aderia, silvicult_ 13068.1263	18 	
	13 23	2015 II Agriculture, gan	aderia, silvicult	4	
	14 24	2015 III Agricultura, gan	aderia, silvicult	<i>a</i>	
	15 18	2015 IV Agricultura, gan	.aderia, milvioult 11407.2164.		
	15	JUIN TOTAL PERIODO Agricultura, gen	Aderia, Bliviolit_ 48500.3		
	10 10	2016 TT Agriculture, gen	solaria, alluinolt. 11940.0018		
	10 72	3016 TTT APPENDIX ON	aderia elletoria 11544.5770		
	20 23	2016 IV Asrivultura, pan	aderia, silvioult		
	21 23	2017 Total Periodo Agricultura, pan	aderia, allvicult. S1051.6685		
	22 22	2027 1 Agriculture, gen	aderia, milwicula15927.0063	ue .	
	23 23	2017 II Agricultura, gan	aderia, silvicult_ 12554.8841		
	24 24	2017 112 Agriculture, gen	aderia, alleicola	10	
	25 28	2017 IV Agricultura, gan	aderia, silvicolt12632.0521		
	26 26	2028 Total Periodo Agricultura, gan	aderia, silwicula_ \$2048.0647	9	
	27 27	2018 I Agricultura, gan	aderia, allvioult_ 14244.3355	4	
	28 28	2018 II Agriculture, gen	aderia, ellvíoult. 12668.3030	16	
	29 2.8	2018 111 Agricultura, gan	aderia, #ilvicult	a	
	30 20	2018 IV Agriculture, gen	aderia, mileioult 12743.99281	4	17251

Figura 8. **Power Query**

Fuente: elaboración propia, empleando Microsoft Power Bi.

En el caso de realizar el reemplazo de un valor, basta con seleccionar un campo, clic izquierdo y seleccionar la opción de reemplazar valores. Al aparecer una ventana, colocamos el valor que deseamos encontrar para reemplazar y posterior el valor por el que vamos a reemplazar. Es importante en opciones avanzadas seleccionar la opción de coincidir con la celda entera, para que solamente sea afectado el valor en especifico.



Figura 9. Reemplazar en Power Query

Fuente: elaboración propia, empleando Microsoft Power Bi.

Finalmente, al realizar nuestras transformaciones automáticas, ya sea algún reemplazo, filtrado o validación, se procede con cerrar y aplicar, y los cambios seran aplicados.

4.1.2.2. Sumarización de resultados

La sumarización nos ayuda a tener una vista resumida de la información mas relevante realizada por medio de DAX y colocando los datos mas relevantes. Para realizar dicho proceso, nos colocamos el menú de Herramientas de Tablas, y crear una nueva tabla.

Figura 10. Creación de tabla en Microsoft Power Bi

			Untitled - Power BI Desktop	
File	He	me Help	Table tools Column tools	
7 Na	e Gere	wal		1
			Mark as date Manage New	Quick New New
			table - relationships measure	measure column table
	Str	udure	Calendars Relationships	Calculations
61	XN	/		
. 1		lo 💌 Periodo	Categoría	 Valor Q(M)
J	1	2013 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	43121.567406
	6	2014 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	45065.826967
	22	2015 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	47445.309669
	26	2026 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	48500.389997
	21	2017 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	\$1051.668507
	26	2018 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	52048.054727
	32	2019 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	55926.03602
	36	2020 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	61382.2376589172
	42	2021 Total Per	odo Agricultura, ganaderia, silvicultura y pesca	33297.8248877038
	44	2012 Total Per	odo Explotación de minas y canteras	5826.65729
	49	2014 Total Per	odo Explotación de minas y canteras	7372.692882
	54	2015 Total Per	odo Explotación de minas y canteras	5571.744423
	59	2016 Total Per	odo Explotación de minas y canteras	5076.53406
	64	2017 Total Per	odo Explotación de minas y canteras	3743.235362
	49	2018 Total Per	odo Explotación de minas y canteras	2982.755159
	74	2019 Total Per	odo Explotación de minas y canteras	3145.221634
	79	2020 Total Per	odo Explotación de minas y canteras	3031.25832394643
	.84	2021 Total Per	odo Explotación de minas y canteras	1865-58808194714
	87	2013 Total Per	odo Industrias manufactureras	60027.565965
	92	2014 Total Per	odo Industrias manufactureras	62962.556965
	57	2015 Total Per	odo Industrias manufactureras	68196.593593
	302	2016 Total Per	odo Industrias manufactureras	71237.057350
	\$07	2017 Total Per	odo industrias manufactureras	74450.475933
	222	2018 Total Per	odo Industrias manufactureras	76870.290094
	337	2019 Total Per	odo Industrias manufactureras	\$1910.716349
	222	2020 Total Per	odo Industrias manufactureras	84430.4540400812
	227	2021 Total Per	odo Industrias manufactureras	45122.9582016808
	130	2013 Total Per	odo Suministro de electricidad, agua y saneamiento	11882.6253376483
	235	2014 Total Per	odo Summstro de electricidad, agua y saneamiento	12541.2592514623
	340	2015 Total Per	odo Suministro de electricidad, agua y saneamiento	11201.0526750011
	245	2016 Total Per	odo Suministro de electricidad, agua y saneamiento	11913.6564741246
	250	2017 Total Per	odo Suministro de electricidad, agua y sameamiento	12021.2262733766
	155	2018 Total Per	odo Suministro de electricidad, agua y saneamiento	12086.3672283847
	260	2019 Total Per	odo Summstro de electricidad, agua y saneamiento	23846.3326362773
	365	2020 Total Per	odo Suministro de electricidad, agua y saneamiento	14101.6392489660
	270	2021 Total Per	odo Suministro de electricidad, agua y saneamiento	7247.25587929256
	273	2013 Total Per	odo Construcción	17835.3793154056
	178	2014 Total Per	ada Construcción	10064 5767307002

Fuente: elaboración propia, empleando Microsoft Power Bi.

Por medio de DAX, se realizará la creación de la tabla, colocando primero el nombre de la tabla, seguido del signo igual, que es donde se realizara el manejo de la información. En este caso, se tomará de la hoja General la información para mostrar la suma de categorías por años.

Figura 11. Creación de sumarización por DAX



Fuente: elaboración propia, empleando Microsoft Power Bi.

Finalmente, se acepta la creación de la tabla, y se observara la nueva tabla creada, junto con la información sumarizada.

Figura 12. Tabla sumarizada

8	୨୯ ୴	ititled - Power BI Desktop	₽ Search	Sprin 🔍 🗕 Of 🗙
File	Home Help Table tools			
Ø Nat	The DatosSumarizados Mark as date table ~ Structure Calendars	Manage relationships Relationships Calcular	New New Column table 995	
bd.	X V 1 DetosSumerizados - SUMM	WAITE(L v Eislete >>
	2 FILTER("General", MOT("	General'[Periodo] in ("Total P	eriodo"))),General[Año],General[Categoria],"Montos Totales",SUM(General[Valor Q(M)]))	E Paus '
an i	Categoria	Alio 🔻 Montos Yotales 💌		P Search
	gricultura, ganaderia, silvicultura y pesca	2013 43121.567406		
42 ×	gricultura, ganaderia, silvicultura y pesca	2014 45065.826967		> I DatosSumarizados
	gricultura, ganaderia, silvicultura y pesca	2015 47445.209669		7 III ORINA
	gricultura, ganaderia, silvicultura y pesca	2016 48500.369997		> 100 5
	gricultura, ganaderia, silvicultura y pesca	2017 51051.668507		
	gricultura, ganaderia, silvicultura y pesca	2018 52048.064717		
	gricultura, ganaderia, silvicultura y pesca	2019 55926.02602		
	gricultura, ganaderia, silvicultura y pesca	2020 61362.2376989272		
	gricultura, ganaderia, silvicultura y pesca	2021 33297.8148877018		
	igilotación de minas y canteras	2013 5826.65729		
	igilotación de minas y canteras	2014 7372.692882		
	igilotación de minas y canteras	2015 5571.744433		
	iplotación de minas y canteras	2016 5076.53406		
	igilitación de minas y canteras	2017 \$743.135362		
t (iplotación de minas y canteras	2018 2983.755259		
	iplicitación de minas y canteras	2019 2145.221634		
	igliotación de minas y canteras	2020 3031.25832254643		
1	igilotación de minas y canteras	2021 1865.55508254724		
	idustrias manufactureras	2013 60027.565965		
	idustrias manufactureras	2014 62968.556965		
	idustrias manufactureras	2015 64196.530590		
	idustrias manufactureras	2016 71237.057898		
	idustrias manufactureras	2017 74450.475933		
	idustrias manufactureras	2018 76870.290094		
	idustrias manufactureras	2019 81818.726349		
	idustrias manufactureras	2020 84430.4540400822		
	idustrias manufactureras	2021 45122.9982016808		
3	uministro de electricidad, agua y saneamiento	2013 11882.6253376483		
5	uministro de electricidad, agua y saneamiento	2014 12541.2992514622		
5	uministro de electricidad, agua y saneamiento	2015 11281.8526758011		
5	uministro de electricidad, agua y saneamiento	2016 11913.6564742246		
5	uministro de electricidad, agua y saneamiento	2017 12021.2261783766		
3	uministro de electricidad, agua y saneamiento	2018 12086.3672283847		
5	uministro de electricidad, agua y saneamiento	2019 11046.3516361772		
5	uministro de electricidad, agua y saneamiento	2020 14181.6392489668		
3	uministro de electricidad, agua y saneamiento	2021 7247.25587929256		
0	anetrucción	2013 17835.3793154056		
	pretrucción	2014 20864.5757507883		×
SOF D1	IDSSUMARZADOS (152 IÓWS)			

Fuente: elaboración propia, empleando Microsoft Power Bi.
5. ANALÍTICA DE DATOS

5.1. Análisis estadístico descriptivo

Por medio de este análisis tenemos un resumen de la información que dan los datos de una muestra para verificar la precisión, sencillez y la calidad de los datos. Logrando los siguientes objetivos:

- Recolectar y ordenar la información
- Extraer las características mas representativas de una colección de datos.
- Describir las tendencias.

5.1.1. Sumario de estadística descriptiva total país

El análisis realizado por medio de R es el resumen del tota de los datos del PIB del país desde el año 2013 hasta el segundo trimestre del 2021, los datos analizados fueron el dato mas grande, el mas pequeño, media, mediana, primer cuartil y tercer cuartil por categoría en Q(M), ya que este es el campo para analizar. Dicho análisis puede ser observado en la tabla I.

Tabla I. Sumario estadístico total pais en millones de quetzales

	Min.	Max.	Mediana	Media	1er	3er
					Cuartii	Cuartii
Total del país	591,2	31 429,9	5 127,7	7 242,2	3 682,1	8 199,4

Fuente: elaboración propia, empleando Microsoft Power Bi.

5.1.2. Sumario de estadística descriptiva por categoría

El análisis realizado por medio de R es el resumen de los datos agrupado por las categorías del PIB de Guatemala desde el año 2013 hasta el segundo trimestre del 2021, la información que se presenta es dato mas grande, el mas pequeño, media, mediana, primer cuartil y tercer cuartil por categoría en Q(M), ya que este es el campo para analizar. Dicho análisis puede ser observado en la tabla II.

Tabla II. Sumario estadístico por categoría en millones de quetzales

Categoria	Min. Q(M)	Max. Q(M)	Mediana Q(M)	Media Q(M)	1er cuartil Q(M)	3er cuartil Q(M)
Actividades de alojamiento y de servicio de comidas	2 539	5 209	3 875	3 765	3 202	42 666
Actividades de servicios administrativos y de apoyo	3 355	5 555	4 287	4 349	4 024	4 624
Actividades financieras y de seguros	3 397	6 202	4 652	4 761	4 645	5 557
Actividades inmobiliarias	8 867	13 500	10 934	11 031	9 881	12 181
Actividades profesionales científicas y técnicas	2 420	4 048	3 113	3 153	2 839	3 440
Transporte y almacenamiento	2 644	5 463	3 973	3 867	3 498	4 284

Continuación de la tabla II.

Categoria	Min. Q(M)	Max. Q(M)	Mediana Q(M)	Media Q(M)	1er cuartil Q(M)	3er cuartil Q(M)
Comercio y reparación de vehículos	18 568	31 430	24 597	24 642	21 377	27 836
Construcción	4 242	8 767	6 371	6 336	5 359	7 062
Enseñanza	4 085	8 497	5 938	5 987	5 108	6 740
Explotación de minas y canteras	591,2	2 089,7	1 140,4	1 135,8	788,3	1 374,8
Impuestos netos de subvenciones a los productos	5 981	10 087	7 592	7 745	6 822	8 747
Industrias manufactureras	14 226	22 922	7 592	7 745	6 822	8 747
Información y comunicaciones	4 434	6 074	5 228	5 305	5 137	5 581
Otras actividades de servicios	3 942	6 297	5 266	5 210	4 638	5 759
Salud	2 024	5 113	3 183	3 257	2 747	3 267
Suministro de electricidad, agua y saneamiento	2 487	3 925	3 097	3 147	2 940	3 291
Administración pública y defensa	3 717	7 798	5 369	5 404	4 585	6 055
Agricultura, ganaderia, silvicultura y pesca	9 920	17 114	12 594	12 878	11 592	13 999

Fuente: elaboración propia, empleando RStudio.

5.1.3. Análisis de datos explotatorios y análisis de *outliers* por categoría

El análisis de datos exploratorio (EDA) es una parte importante en el análisis de datos, ya que nos ayuda a determinar la calidad, limpieza y expectativas que deben cumplir los datos. Por medio de este análisis se puede determinar que la fuente de datos haya sido obtenida sin problema, ya que, si este análisis corresponde a un mal resultado, se deben tomar de nuevo los datos, porque no brindarían un resultado confiable. Para cada categoría se tienen los siguientes resultados:

 Comercio y reparacion de vehiculos: como se puede observar en la figura 13, en el histograma vemos en los rangos de 18 000 Q(M) a 20 000 Q(M) y de 24 000 Q(M) a 26 000 Q(M) tenemos la mayor concentración de valores. En el *boxplot* podemos notar que el rango de datos limpios significativos está en el rango de 20 000 Q(M) a 28 000 Q(M) teniendo su media en 24 000 Q(M). Finalmente, según la curva de tendencia podemos confirmar que los valores en el rango de 20 000 Q(M) a 28 000 Q(M) son los que mejor se ajustan y generan menos ruido. en la figura 14, se puede observar que el valor medio de Comercio y Reparacion de Vehiculos, presenta desde el año 2013 un crecimiento hacia el año 2019.



Fuente: elaboración propia, empleando RStudio.





Fuente: elaboración propia, empleando RStudio.

Agricultura, ganadería, silvicultura y pesca: se puede observar en la figura 15, en el histograma vemos en los rangos de 11 000 Q(M) a 12 000 Q(M) tenemos la mayor concentración de valores. En el *boxplot* podemos notar que el rango de datos limpios significativos está en el rango de 11 000 Q(M) a 15 000 Q(M) teniendo su media en 12 500 Q(M). Finalmente, según la curva de tendencia podemos confirmar que los valores en el rango de 11 000 Q(M) a 15 000 Q(M) son los que mejor se ajustan y generan menos ruido. En la figura 16, se puede observar que el valor medio de Agricultura, ganadería, silvicultura y pesca desde el año 2013 hasta el año 2019 no presento un incremento significativo en la media de sus valores. Teniendo hasta el año 2020 y lo que va del 2021 que la tendencia de la media es un crecimiento considerable.

Figura 15. EDA de agricultura, ganadería, silvicultura y pesca



Fuente: elaboración propia, empleando RStudio.



Figura 16. *Outlier* de agricultura, ganadería, silvicultura y pesca

Fuente: elaboración propia, empleando RStudio.

Explotación de minas y canteras: se puede observar en la figura 17, en el histograma vemos en los rangos de 600 Q(M) a 700 Q(M) tenemos la mayor concentración de valores, y que los valores se encuentran distribuidos en varios rangos. En el *boxplot* podemos notar que el rango de datos limpios significativos está en el rango de 500 Q(M) a 1 400 Q(M) teniendo su media en 1 200 Q(M); mientras que la curva de tendencia podemos notar que el rango es de 500 Q(M) a 1 700 Q(M), un rango mas amplio a comparación del boxplot. En la figura 17, se puede observar que el valor medio de Explotación de minas y canteras tuvo su mayor auge en el 2014, pero en los años posteriores fue en caída la media, a partir del año 2019 tuvo un pequeño incremento no signicativo, con lo que estos cambios radicales justifican el por qué con el EDA, los rangos de datos limpios son variantes.



Figura 17. EDA de explotación de minas y canteras

Fuente: elaboración propia, empleando RStudio.



Ingresos vs Años



Fuente: elaboración propia, empleando RStudio.

Industrias manufactureras: se puede observar en la figura 19, en el histograma vemos en los rangos de 19 000 Q(M) a 20 000 Q(M) tenemos la mayor concentración de valores, aunque tiene en los demás rangos se mantienen cercanos y con una tendencia similar. En el *boxplot* podemos notar que el rango de datos limpios significativos está en el rango de 16 000 Q(M) a 20 000 Q(M) teniendo su media en 18 000 Q(M). Finalmente, según la curva de tendencia podemos confirmar que los valores en el rango de 16 000 Q(M) a 20 000 Q(M) a 20 000 Q(M) son los que mejor se ajustan y generan menos ruido. En la figura 20, se puede observar que el valor medio de Industrias manufactureras desde ha mantenido un crecimiento estable y notorio.





Fuente: elaboración propia, empleando RStudio.

Figura 20. Outlier de industrias manufactureras



Fuente: elaboración propia, empleando RStudio.

Suministro de electricidad, agua y saneamiento (SEAS): se puede observar en la figura 21, en el histograma vemos en los rangos 3 000 Q(M) a 3 200 Q(M) se encuentra la mayor frecuencia de datos y que la distribución de los rangos es menos extensa a comparación de las otras categorías. En el *boxplot* podemos notar que el rango de datos limpios significativos está en el rango de 2 900 Q(M) a 3 300 Q(M) teniendo su media en 3 100 Q(M). Finalmente, según la curva de tendencia podemos confirmar que los valores en el rango de 2 900 Q(M) a 3 300 Q(M) son los que mejor se ajustan y generan menos ruido. En la figura 20, se puede observar que el valor medio de Suministro de electricidad, agua y saneamiento del 2014 al 2018 tuvo un comportamiento estable hasta el año 2019 que la tendencia ha sido de crecimiento para la media de esta categoría y se ha mantenido.





Fuente: elaboración propia, empleando RStudio.

Figura 22. Outlier de suministro de electricidad, agua y saneamiento



Fuente: elaboración propia, empleando RStudio.

Construcción: se puede observar en la figura 23, que la mayor frecuencia de los datos esta en el rango de 6 000 Q(M) a 6 200 Q(M). En el *boxplot* el rango de los datos limpios significativos es de 5 500 Q(M) a 7 000 Q(M) teniendo su media en 6 200 Q(M), además que la caja es asimétrica teniendo mas larga la parte inferior, por lo que la mayor parte de datos se concentran en la parte superior de los datos, tal como se ve reflejado en el histograma. Finalmente, según la curva de tendencia podemos confirmar que los valores en el rango de 4 000 Q(M) a 9 000 Q(M) son los que mejor se ajustan y generan menos ruido. En la figura 24, se puede observar que de 2013 a 2014 tuvo un incremento, luego mantuvo estabilidad en el crecimiento hasta el 2019 que tuvo un mayor crecimiento, pero en los próximos años con una tendencia de mantenerse estable.



Figura 23. EDA de Construcción

Fuente: elaboración propia, realizado en RStudio.

Figura 24. Outlier de construcción



Ingresos vs Años

Fuente: elaboración propia, empleando RStudio.

Transporte y almacenamiento: en la figura 25, el histograma muestra que en el rango de 4 000 Q(M) a 4 500 Q(M) se encuentra la mayor concentración de los datos, teniendo muy cercano el rango de 3 500 Q(M) a 4 000 Q(M), siendo estos los que poseen la mayor concentración de los datos que además en el grafico de densidad se confirma donde se encuentra la mayor distribucion. En el *boxplot* se puede observar que la distribución es asimétrica negativa, al tener la media en el valor de 4 000. Por ultimo, en la curva de se observa que del rango de 4 000 Q(M) a 4 500 Q(M) son los datos que mejor se ajustan y generan menor ruido. En la figura 26, se puede observar que a lo largo de los años esta categoría ha tenido un crecimiento en sus valores y el crecimiento por cada año no ha tenido rangos muy extensos a excepción del año 2020, que se puede observar una caja mas grande, lo que significa un rango mayor.



Figura 25. EDA de transporte y almacenamiento

Fuente: elaboración propia, empleando RStudio.





Fuente: elaboración propia, empleando RStudio.

Actividades de alojamiento y de servicio de comidas (AASC): en la figura 27, el histograma muestra que la mayor concentración de los datos se encuentra en el rango de 4 000 Q(M) a 4 500 Q(M), pero en este caso el segundo mayor rango, se encuentra alejado del mayor, por lo que en el grafico de densidad se observa como se encuentra la distribución. En el *boxplot* se observar una asimetría negativa, con lo que la media es menor a la mediana. Mientras que en la curva de tendencia los datos que generan menor ruido y se ajusta mejor están en rango de 2 600 Q(M) a 4 500 Q(M). En la figura 28, se observa que esta categoría mantenía un crecimiento constante hasta el año 2019, siendo en el 2020 cuando tuvo un decrecimiento, pero con se nota una recuperación en el año 2021.





Fuente: elaboración propia, empleando RStudio.

Figura 28. Outlier de actividades de alojamiento y servicio de comidas



Fuente: elaboración propia, empleando RStudio.

Información y comunicaciones: en la figura 29, en el histograma se observa que el rango con mayor frecuencia es de 5 000 Q(M) a 5 250 Q(M), y además que en los demás rangos se encuentran distribuidas las frecuencias de manera similar, y que la distribucion de los datos no es muy grande, en la grafica de densidad se puede confirmar la distribucion mencionada anteriormente. Mientras que en el *boxplot* se puede observar la asimetría positiva por lo que la media de estos datos es mayor que la mediana. Finalmente, la línea de tendencia muestra que los valores que mejor se ajustan están entre en rango de 5 000 Q(M) a 600 Q(M). En la figura 30, se observa que la categoría ha tenido una tendencia de crecimiento cada tres años, y manteniendo en su mayoría una simetría positiva.





Fuente: elaboración propia, empleando RStudio.





Fuente: elaboración propia, empleando RStudio.

Actividades financieras y de seguros: en la figura 31, se observa que en esta categoría se encuentra una distribucion de las frecuencias muy similar entre todos los rangos, aún así siendo la mayor la del rango de 4 000 Q(M) a 4 500 Q(M), y en la grafica de densidad se puede ver reflejado la similutd en la distribucion de las frecuencias en los rangos. Mientras que con el *boxplot* se puede observar que se encuentra con una simetría positiva por lo que se tiene una media mayor a la mediana. Finalmente, en la línea de tendencia se observa que los datos que mejor se ajustan están en el rango de 4 000 Q(M) a 6 000 Q(M). En la figura 32, se observa el crecimiento constante que ha tenido este sector, que cabe destacar que el 2020 ha sido el único año con mayor variación en los datos.

Figura 31. EDA de actividades financieras y de seguros



Fuente: elaboración propia, empleando RStudio.

Figura 32. Outlier de actividades financieras y de seguros



Ingresos vs Años

Fuente: elaboración propia, empleando RStudio.

Actividades inmobiliarias: en la figura 33, se observa un histograma que contiene frecuencias distribuidas de forma muy similar, teniendo sobresaliente el rango de 9 500 Q(M) a 10 000 Q(M), y con el grafico de densidad se puede confirmar. Mientras tanto en el *boxplot* vemos que casi es simétrica la figura, pero por tener el lado superior más amplio se cumple de ser una asimetría positiva dando como resultado una media mayor que la mediana. En el gráfico de la línea de tendencia, se observa que los datos que mejor se ajustan están en el rango de 9 000 Q(M) a 13 000 Q(M) que coinciden con la mayor concentración de frecuencias en el histograma. En la figura 34, se observa que esta categoría ha tenido un crecimiento constante, en donde muchas cajas son simétricas, y que la tendencia se mantiene a seguir creciendo.



Figura 33. EDA de actividades inmobiliarias

Fuente: elaboración propia, empleando RStudio.

Figura 34. *Outlier* de actividades inmobiliarias

Ingresos vs Años



Fuente: elaboración propia, empleando RStudio.

Actividades profesionales centificas y técnicas (APCT): en la figura 35, se observa en el histograma que el rango con la mayor frecuencia es de 2 800 Q(M) a 3 000 Q(M), en el gráfico de densidad, se puede observar como es que en esta categoría se encuentra una diferencia entre la frecuencia máxima y el resto. El *boxplot* se observa que es asimetrico positivo, por lo que la media es mayor que la mediana. Finalmente, en el grafico de línea de tendencia se observar que los valores que mejor se ajustan están en el rango de 2 600 Q(M) a 3 500 Q(M). En la figura 36, se observa un crecimiento irreglar a lo largo del tiempo en esta categoría, manteniendo una estabilidad a partir del año 2018, donde se disminuye considerablemente la tendencia de crecimiento que poseía.





Fuente: elaboración propia, empleando RStudio.

Figura 36. *Outlier* de actividades profesionales, científicas y técnicas



Fuente: elaboración propia, empleando RStudio.

Actividades de servicios administrativos y de apoyo (ASAA): en la figura 37 se puede observar un rango reducido entre las cantidades del histograma, tenienod la mayor frecuencia de datos en el rango de 4 000 Q(M) a 4 500 Q(M), además que en gráfico de densidad se observa una curva casi simétrica ya que los datos se encuentran distribuidos casi uniformemente. Mientras que con el *boxplot* se observa una simetría positiva por lo que la media es mayor a la mediana. Finalmente, en la línea de tendencia se observa que los datos que mejor se ajustan están en el rango de 3 600 Q(M) a 4 800 Q(M). En la figura 38 se observa que a lo largo de los años el crecimiento de esta categoría ha sido moderado, sin existir gran diferencia entre cada año, a excepción del 2021 que se prevee un crecimiento.

Figura 37. EDA de actividades de servicios administrativos y de apoyo



Fuente: elaboración propia, empleando RStudio.

Figura 38. Outlier de actividades de servicios administrativos y de

ароуо



Ingresos vs Años

Fuente: elaboración propia, empleando RStudio.

Administración publica y defensa: en la figura 39, se observa en el histograma un rango amplio, teniendo la frecuencia más amplia en los rangos de 5 000 Q(M) a 5 500 Q(M), en el grafico de densidad se ve reflejado la distribucion de las frecuencias. En el *boxplot* se observa que esta categoría es asimétrica negatvia teniendo la mediana mayor a la media. Finalmente, en el grafico de tendencia se observa que los datos que se mejor se ajustan están en el rango de 4 000 Q(M) a 7 000 Q(M). En la figura 40 se observa que a lo largo de los años se ha tenido un crecimiento constante en lo que se ha invertido en esta categoría, por lo que se puede deducir que cada año el gobierno inviernte más en en esta categoría.





Fuente: elaboración propia, empleando RStudio.

Figura 40. Outlier de administración pública y defensa



Ingresos vs Años

Fuente: elaboración propia, empleando RStudio.

Enseñanza: en la figura 41, se puede observar que enseñanza se encuentra una distribución similar en los diversos rangos, con la excepción de la mayor frecuencia que es en el rango de 5 000 Q(M) a 5 500 Q(M), en la densidad se observa reflejado. Mientras que en el *boxplot* se puede observar que es simétrico por lo que se tiene que la moda, la mediana y media coinciden. Mientras que en la línea de tendencia se observa que los valores que mejor se ajustan están en el rango de 4 500 Q(M) a 6 500 Q(M). En la figura 42 se observa un crecimiento constante a lo largo de los años.



EDA de enseñanza



Fuente: elaboración propia, empleando RStudio.



Outlier de enseñanza



Fuente: elaboración propia, empleando RStudio.

Salud: en la figura 43, se observa que el histograma de salud tiene la mayor parte de sus frecuencias acumuladas en los rangos desde 2 000 Q(M) hasta 3 500 Q(M), que se refleja en el gráfico de densidad. Mientras que el boxplot se observa simétrico, por lo que la moda, mediana y media coinciden. Finalmente, la línea de tendencia se observa que los valores que mejor coinciden se encuentran en el rango de 2 500 Q(M) a 3 500 Q(M). En la figura 44, se observa un crecimiento constante a lo largo de los años, por lo que se puede deducir que cada año la inversión de salud ha ido incrementando, sobre todo en 2020, que, por situación de pandemia, el gasto de salud tuvo que ser incrementado notablemente.



Figura 43. EDA de salud

Fuente: elaboración propia, empleando RStudio.

Figura 44. Outlier de salud





Fuente: elaboración propia, empleando RStudio.

5.2. Análisis predictivo

Para el análisis predtictivo se uso el método de forecasting ARIMA, por ser el modelo mas completo ya que toma en cuenta desde la parte autogresiva hasta el termino de estacionariedad. Por otra parte, se actualizó el período de la información del banco de Guatemala hasta el cuarto trimestre del año 2021, para que se tenga un análisis predictivo más asertivo en cuanto a la situación actual del país.

5.2.1. Análisis predictivo del total del país

Se realizó un análisis predictivo tomando en cuenta el segundo y tercer trimestre del año 2020 y uno sin estos trimestres, debido a que por efectos de pandemia

de COVID-19, durante estos periodos de tiempo es cuando el país se vio mas afectado. Como se observa en la figura 45, para el periodo de tiempo en que transcurrió la pandemia, existe un pico de caía de las actividades, y esta refleja en la predicción que ocurriría este mismo momento, pero el covid al ser un evento extraordinario, no puede tomarse para el análisis. Por otra parte, en la figura 46, se observa que al quitar esta caída que causó el COVID-19, se tiene un análisis predictivo más certero, por lo que, para las demás categorías del PIB, se tomaron en cuenta los datos excluyendo segundo y tercer trimestre del año 2020.



Figura 45. Análisis predictivo total país con COVID-19

Fuente: elaboración propia, empleando Microsoft Power Bi.



Figura 46. Análisis predictivo total país sin COVID-19

Fuente: elaboración propia, empleando Microsoft Power Bi.

5.2.2. Análisis predictivo por categoría

A continuación, se presenta el análisis predictivo realizado por cada categoría del PIB de Guatemala:

 Actividades de alojamiento y de servicio de comidas: en la tendencia mostrada en la figura 47, se observa que para las Actividades de alojamiento y de servicio de comidas, se generará una leve alza (marcada por la línea naranja) al igual que en el Q3 del 2020, siendo este su pico máximo en 7 años. Las sombras de error pueden identificarse como impactos del relajamiento o endurecimiento de las medidas gubernativas del país a causa del Covid y sus variantes.

Figura 47. Análisis predictivo actividades de alojamiento y de servicio de comidas



Fuente: elaboración propia, empleando Microsoft Power Bi.

Actividades de servicios administrativos y de apoyo: en la tendencia mostrada en la figura 48, se observa que para Actividades de servicios administrativos y de apoyo, se generará un alza en los últimos 3 años de pronóstico a un ritmo (pendiente de la gráfica) menor como se venía del 2013 al 2021 pero superando el pico más alto ocurrido en el 2020. Se denotan las sombras de error oscilan entre 5,57 Q(MM) como el punto más bajo y 6,97 Q(MM) como el punto más alto.

Figura 48. Análisis predictivo actividades de servicios administrativos y de apoyo



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Actividades financieras y de seguros: en la figura 49, se observa para Actividades financieras y de seguros un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada. El pico máximo se da en el último año donde las sombras de error oscilan 6,75 Q(MM) y 8,11 Q(MM).

Figura 49. Análisis predictivo actividades financieras y de seguros



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Actividades inmobiliarias: en la figura 50, se observa un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada, el ritmo (pendiente de la gráfica) del pronóstico se muestra mayor en comparación de los años 2013 a 2020. El pico máximo se da en el último año donde las sombras de error oscilan entre 13,95 Q(MM) y 17,64 Q(MM).





Fuente: elaboración propia, empleando Microsoft Power Bi.

Actividades profesionales científicas y técnicas: en la figura 51, se observa que para Actividades profesionales científicas y técnicas una tendencia de crecimiento superior a los 8 años previos, aunque el pronóstico tiene picos de subida y bajadas, en la última fecha pronosticada se observa un crecimiento superior en comparación a la tendencia de los años 2013 a 2020. Se denotan que las sombras de error oscilan entre 4,15 Q(MM) como el punto más bajo y 5,10 Q(MM) como el punto más alto.

Figura 51. Análisis predictivo actividades profesionales científicas y técnicas



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Administración pública y defensa: en la tendencia mostrada en la figura 52, se observa que, para las Administración pública y defensa, se generará una leve alza (marcada por la línea naranja) superando por poco al de Q1 del 2021, siendo este su pico máximo en 7 años. El pico máximo se da en el último año donde las sombras de error oscilan entre 5,36 Q(MM) y 10,15 Q(MM).





Fuente: elaboración propia, empleando Microsoft Power Bi.

 Agricultura, ganadería, silvicultura y pesca: en la tendencia observada en la figura 53, se observa que, para Agricultura, ganadería, silvicultura y pesca, se generará un crecimiento con picos de subida y bajada, provocado por las temporadas de cosecha donde se obtienen las mayores ventas, sin embargo, el crecimiento no es tan variante a comparación del pico más alto alcanzando en Q1 de 2021. Se denotan que las sombras de error oscilan entre 13,55 Q(MM) como el punto más bajo y 22,79 Q(MM) como el punto más alto.
Figura 53. Análisis predictivo agricultura, ganadería, silvicultura y pesca



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Comercio y reparación de vehículos: en la figura 54, se observa para Comercio y reparación de vehículos un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada. Se denotan que las sombras de error oscilan entre 34,78 Q(MM) como el punto más bajo y 54,27 Q(MM) como el punto más alto.

Figura 54. Análisis predictivo comercio y reparación de vehículos



Fuente: elaboración propia, empleando Microsoft Power Bi.

Construcción: en la figura 55, se observa para Construcción una leve caída (marcada por la línea anaranjada) para el año 2022, pero en los próximos años el ritmo (pendiente de la gráfica) muestra un crecimiento, llegando a su punto máximo superar el el pico más alto por la tendencia de los pasados 8 años. El pico máximo se da en el último año donde las sombras de error oscilan entre 7,76 Q(MM) y 12,39 Q(MM).

Figura 55. Análisis predictivo de construcción



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Enseñanza: en la figura 56, se observa que para Enseñanza una tendencia de crecimiento superior a los 8 años previos, aunque el pronóstico tiene picos de subida y bajadas, en la última fecha pronosticada se observa un crecimiento superior en comparación a la tendencia de los años 2013 a 2020. Se denotan que las sombras de error oscilan entre 6,61 Q(MM) como el punto más bajo y 10,06 Q(MM) como el punto más alto.



Figura 56. Análisis predictivo de enseñanza

Fuente: elaboración propia, empleando Microsoft Power Bi.

Explotación de minas y canteras: en la figura 57, se observa que la proyección es plana como se observa en el año 2021, sin embargo, las sombras de error oscilan entre los -0,10 Q(MM) y 2,03 Q(MM). Además, debido a las leyes del país con respecto a la minería es que se observa el comportamiento de la caída en esta categoría.



Figura 57. Análisis predictivo de explotación de minas y canteras

Fuente: elaboración propia, empleando Microsoft Power Bi.

 Industrias manufactureras: en la figura 58, se observa para Industrias manufactureras un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada. Se denotan que las sombras de error oscilan entre 21,99 Q(MM) como el punto más bajo y 28 Q(MM) como el punto más alto.





Fuente: elaboración propia, empleando Microsoft Power Bi.

Información y comunicaciones: en la tendencia mostrada en la figura 59, se observa que, para las Información y comunicaciones, se generará una leve alza (marcada por la línea naranja) superando por poco al de Q2 del 2021, siendo este su pico máximo en 7 años. El pico máximo se da en el último año donde las sombras de error oscilan entre 5,52 Q(MM) y 7,35 Q(MM).



Figura 59. Análisis predictivo de información y comunicaciones

Fuente: elaboración propia, empleando Microsoft Power Bi.

 Salud: en la figura 60, se observa Salud un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada. Se denotan que las sombras de error oscilan entre 34,78 Q(MM) como el punto más bajo y 54,27 Q(MM) como el punto más alt.





Fuente: elaboración propia, empleando Microsoft Power Bi.

 Suministro de electricidad, agua y saneamiento: en la figura 61, se observa para Suministro de electricidad, agua y saneamiento que la proyección es plana como se observa en el año 2021, sin embargo, las sombras de error oscilan entre los 2,46 Q(MM) y 5,57 Q(MM).

Figura 61. Análisis predictivo de suministro de electricidad, agua y saneamiento



Fuente: elaboración propia, empleando Microsoft Power Bi.

 Transporte y almacenamiento: en la figura 62, se observa en Transporte y almacenamiento un crecimiento superior a los últimos 8 años, previo a la tendencia mostrada. Se denotan que las sombras de error oscilan entre 4,86 Q(MM) como el punto más bajo y 6,79 Q(MM) como el punto más alto.





Fuente: elaboración propia, empleando Microsoft Power Bi.

6. CREACIÓN DE TABLERO DE CONTROL

6.1. Generación de visualizaciones

Para la realización del tablero de control, fue requerido previo el proceso de trasnformación de datos como se muestra en Capitulo 4, una vez realizado ese proceso se procede a la generación de las visualizaciones como se muestra a continuación. Primero debemos observar que nuestro Dataset se encuentra correctamente cargado, como se muestra en la figura 63.

Figura 63. Dataset cargado en Microsoft Power Bi



Fuente: elaboración propia, empleando Microsoft Power Bi.

Posterior a ello, debemos tener instalado Forecasting with Arima, que se puede instalar desde Microsoft Power Bi visuals. Una vez instalado, debe aparecer como se muestra en la figura 64.

Figura 64. Forecasting with Arima



Fuente: elaboración propia, empleando Microsoft Power Bi.

Como siguiente paso, se agrega un slicer, para poder seleccionar la categoría que se desea ver, y a la vez se le agrega el campo de la categoría, como se observa en la figura 65.

Figura 65.

Slicer de categoría



Fuente: elaboración propia, empleando Microsoft Power Bi.

Posterior al paso anterior, agregamos nuestra gráfica de Forecasting con Arima, colocando el valor de Fecha que se tomará para el período de tiempo, y el valor de Q(MM) como dimensional a analizar, tal como se muestra en la figura 66.



Figura 66. Agregando gráfica de forecasting con ARIMA

Fuente: elaboración propia, empleando Microsoft Power Bi.

Finalmente, tendremos las visualizaciones creadas, como se muestra en la figura 67.



Figura 67. Visualizaciones en Microsoft Power Bi

Fuente: elaboración propia, empleando Microsoft Power Bi.

6.2. Publicación de las visualizaciones

Para realizar la publicación de la visualización, primero es necesario crear un espacio de trabajo por medio de Microsoft 365, para que esta esté cargada en la nube, y se pueda tener la visualización. Una vez ingresado a Microsoft 365, se accede a la aplicación de Microsoft Power Bi, como se muestra en la figura 68.

Figura 68. Aplicación de Microsoft Power Bi en Microsoft 365

🏮 Página principal de Microsoft Oli 🗙	👍 Power Bl 🛛 🗙 🕂				~ - a ×
← → C 🔒 app.powerbi.com	n/home			표 & ☆ 🚺 🕹 🇯	🔲 🍮 Incógnito 🗄
III Power BI Inicio				P Bus	car ⁹ 🛞
E Inicio	Buenas noches, Herb	ert Rafael ar y compartir conclusiones contro	iladas por datos.	+ Nuevo ir	nforme 🗮 🗸
 Conjuntos de datos Goals Aplicaciones Más información 	Recomendado	P Introducción a Power BI	Introducción a Power Bi Introducción a Power Bi	P Introducción a Power BI	P Introduc
 □ Áreas de trabajo > (2) Mi área de trabajo ∨ 	Conocimiento de los conceptos bá	Introducción: ¿Qué es Power BI?	Inicio rápido: Navegación por el se	Visualización y comprensión de un	And a second sec
↗ Obtener datos	Recientes Favoritos Mis aplicac	iones			Ver todo



Lo siguiente, es crear un espacio de trabajo, en donde se estarán publicando las visualizaciones, como se muestra en la figura 69.

Figura 69. Creación de Espacio de trabajo



Fuente: elaboración propia, empleando Microsoft Power Bi.

El siguiente paso, es en Microsoft Power Bi de versión de escritorio, irse a la opción de publicar. Posterior a ello, se procede a seleccionar el espacio de trabajo creado anteriormente como se muestra en la figura 70.

Figura 70. Publicación de visualización

85	୍ୟ Tesis - Power Bl Desktop	₽ Search		Herbert Rafael Reyes Portil	•• −
File	Home Insert Modeling View Help				
Paste	Cory Crow Format painter proceed	Cent Transform Refresh Gamerie Wilsel box Wiskly Gamerie Wilsel box Wiskly Centre Wiskly box Wiskly Centre Wiskly box Wiskly Centre Wisk	Dish are		^
<u>Ini</u>	Actividades financi	eras y de seguros	«	Visualizations >>>	Fields >>>
	8-	Solect all Astroitestes de abrives	mianta unda requicio d	Build visual	P Search
6日		Publish to Power Bl	× vos y Iters		> 🎛 DatosSumarizados
	7-	Select a destination	y téon	E la E la E la	> III General
		Search	pesca		□ ∑ Año
	ξ.	My workspace			Categoria
	ar Q()	Tablero PIB Guatemala	los pr	R Py E	D ID
	Val			📲 🖵 🗄 📓 🚼 🛇	Periodo
			aneami	∞ …	Valor Q(MI)
				2	> 🎟 Тор 5
				Values	
		Select Ca	ancel	Add data fields here	
	03/13 06/14 09/15 11/16 02/18	Fecha		Cross-report	
				Keep all filters	
				Add drill-through fields here	
4	Resumen Resumen por año Datos Covid Datos S	n Covid General Sin Covid General Covid 🔶			

Fuente: elaboración propia, empleando Microsoft Power Bi.

Al finalizar la publicación, se podrá observar el tablero cargado como se muestra en la figura 71.



Figura 71. Tablero publicado en web de Microsoft Power Bi

Fuente: elaboración propia, empleando Microsoft Power Bi.

Se logra ver el tablero publicado en la aplicación de Microsoft Power Bi, como se muestra en la figura 72.



Figura 72. Visualizacion de Tablero Publicado

Fuente: elaboración propia, empleando Microsoft Power Bi.

La siguiente parte es publicar el tablero en un sitio, por lo que se utilizará por medio de Microsoft SharePoint. Como se muestra en la figura 73, se procede a la creación de un Microsoft SharePoint.

Figura 73. Creación de Microsoft SharePoint



Fuente: elaboración propia, empleando Microsoft SharePoint.

La siguiente parte consiste en desde la aplicación web de Microsoft Power Bi, seleccionar la opción de archivo, insertar informe y publicar en la web. Como se muestra en la figura 74.



Figura 74. Publicar informe en la web

Fuente: elaboración propia, empleando Microsoft Power Bi.

Al dar la opción de compartir, se procede a copiar el código HTML que puede ser pegado en un sitio web, como se muestra en la figura 75.



Figura 75. Código para publicar en sitio web

Fuente: elaboración propia, empleando Power Bi.

En Microsoft SharePoint, se selecciona la opción de insertar código como se muestra en la figura 76.



Figura 76. Insertar código en Microsoft SharePoint

Fuente: elaboración propia, empleando Microsoft SharePoint.

Una vez seleccionada la opción de insertar código, se pega en la caja de texto nuestro código copiado desde Microsoft Power Bi, como se muestra en la figura 77.



Figura 77. Pegando código de Microsoft Power Bi

Fuente: elaboración propia, empleando Microsoft SharePoint.

Finalmente, se puede observar la gráfica creada en Microsoft SharePoint, como se muestra en la figura 78.



Figura 78. Dashboard publicado en Microsoft SharePoint

Fuente: elaboración propia, empleando Microsoft SharePoint.

CONCLUSIONES

- 1. El uso de la Ciencia de Datos cobra un mayor auge en la actualidad habilitando la recopilación, transformación, análisis y visualización de cualquier información que puede ser usada para la toma de decisiones.
- 2. El Producto Interno Bruto (PIB) en Guatemala, se conforma de varias actividades o categorías económicas, entre las cuales en el 2021 se pronostican tendencias positivas con crecimientos en los próximos 5 años. Algunas otras, muestran tendencias estables o a la baja las cuales será importante razonar qué acciones inmediatas se deben tomar para equilibrarlas.
- Por medio de un análisis estadístico descriptivo (EDA Exploratory Data Analysis), se puede asegurar la calidad de los datos a analizar y como resultado tener un pronóstico que sea más asertivo. Removiendo los outliers (valores atípicos).

RECOMENDACIONES

- 1. Mantener la fuente de datos (*dataset*) del PIB del Banguat actualizado al día para mejorar la calidad de pronósticos en el año actual o consecutivo.
- Actualizar trimestralmente la fuente de datos y las visualizaciones de la herramienta generada para que el usuario final tenga pronósticos de las categorías de interés sobre el PIB de forma actualizada.
- Enriquecer el análisis a través de otros algoritmos de uso de pronósticos más allá del algortimo ARIMA.
- 4. Realizar el respectivo análisis de datos exploratorio (EDA) con la información a ingestar, para garantizar la calidad.
- 5. Remover el ruido causado por el año 2020, ya que solamente Q 1 y Q 4 fueron cercanos a la media, mientras que Q 2 y Q 3 no.

BIBLIOGRAFÍA

- Banco de Guatemala. Producto bruto interno trimestral. [en línea]. <https://banguat.gob.gt/sites/default/files/banguat/cuentasnac/PIB2 013/PDF_graficas_y_cuadros_estadisticos.pdf />. [Consulta: 4 de octubre de 2021].
- Banco Mundial. Crecimiento del PIB (% anual) Guatemala. [en línea]. https://datos.bancomundial.org/indicator/NY.GDP.MKTP.KD.ZG?l ocations=GT >. [Consulta: 4 de octubre de 2021].
- Economipedia. Producto interior bruto (PIB). [en línea].
 https://economipedia.com/definiciones/producto-interior-bruto-pib.html >. [Consulta: 4 de octubre 2021].
- Heavy.AI. Data integration definition. [en línea].
 https://www.heavy.ai/technical-glossary/data-integration.
 [Consulta: 3 de octubre 2021].
- 5. IBM. *Data science*. [en línea]. < https://www.ibm.com/cloud/learn/datascience-introduction >. [Consulta: 27 de septiembre 2021].
- Investopedia. Data analytics. [en línea].
 https://www.investopedia.com/terms/d/data-analytics.asp. [Consulta: 3 de octubre 2021].

- Master's in Data Science. What is data analytics? [en línea].
 https://www.mastersindatascience.org/learning/what-is-data-analytics/. [Consulta: 3 de octubre 2021].
- Microsoft. ¿Qué es microsoft power bi? [en línea].
 https://docs.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>. [Consulta: 2 de octubre 2021].
- Acerca de power query en microsoft excel. [en línea].
 https://support.microsoft.com/es-es/office/acerca-de-power-query-en-Excel-7104fbee-9e62-4cb9-a02e-5bfb1a6c536a
 [Consulta: 2 de octubre 2021].
- Componente regresión lineal. [en línea].
 https://docs.microsoft.com/es-es/azure/machinelearning/component-reference/linear-regression >. [Consulta: 3 de octubre 2021].
- Elegir la mejor línea de tendencia para los datos. [en línea].
 ">https://support.microsoft.com/es-es/office/elegir-la-mejor-l%C3%ADnea-de-tendencia-para-los-datos-1bb3c9e7-0280-45b5-9ab0-d0c93161daa8>. [Consulta: 28 de septiembre 2021].
- R-project. What is R? [en línea]. <https://www.r-project.org/about.html >.
 [Consulta: 28 de septiembre 2021].
- RPubs. Modelos de pronóstico de tendencia y estacionalidad con series temporales. [en línea]. <https://rpubs.com/Ledesma6/586796>. [28 de septiembre 2021].

- RStudio. Take control of your R code. [en línea].
 https://www.rstudio.com/products/rstudio/. [Consulta: 28 de septiembre 2021].
- 15. Significados. Significado de PIB. [en línea]. <https://www.significados.com/pib/>. [Consulta: 4de octubre de 2021].

APÉNDICES

Apéndice 1. Instalación de Librerias de R



Fuente: elaboración propia, empleando RStudio.

Apéndice 2. Creación de dataset en RStudio

sminal × 1005					Environment History Connections Tutorial	
Import Excel Data						
File/URL:						
C:/Users/Her	bert/Desktop/	Datos Tesis xisx			Browse	ε.
Oute Dansieur						-
ID (thuble)	Ato v	Periodo Sheradari v	Categoria Steracteri *	Valor Q(M) (Stuble)		
	2013	Total Periodo	Apricultura, garageria, silvicultura y pesca	43121.567		1
	2013		Agricultura, garaderia, silvicultura y pesca	12146.334		1
1	2013	11	Agricultura, garaderia, silvicultura y pesca	10728.446		1
4	2013	11	Apricultura, garaderia, silvicultura y pesca	9919.570		
1	2013	IV.	Apricultura, garaderia, silvicultura y pesca	10927.217		
	2014	Total Periodo	Apricultura, garaderia, silvicultura y pesca	49065-817		
	2014		Apricultura, garaderia, silvicultura y pesca	11718.791		
	2014	1	Apricultura, garaderia, silvicultura y pesca	10930.561		
1	2014	11	Apricultura, garaderia, silvicultura y pesca	12928-279		
10	2014	IV.	Apricultura, garaderia, silvicultura y pesta	11487.786		
11	20/1	Total Periodo	Apricultura, garaderia, silvicultura y pesca	47445.210		
13	2013		Apricultura, garaderia, silvicultura y pesca	13068-126		
13	2013	11	Apricultura, garaderia, silvicultura y pesca	11738-253		
14	2013		Apricultura, garaderia, silvicultura y pesca	11231.614		
19	2019	IV	Apricultura, garaberla, silvicultura y pesca	11407.216		
1	2019	10131 Periodo	Aproutura, garadera, sivicutura y pesca	48900.370		
	2011		Apriculture, garagere, sinculture y peca	10070-027		
	2011		Andrehen gereicht, strictiste y pool	11040.001		
				110401001		
Previewing first	50 entries.					
Import Option	1				Code Preview	¢
Narre: Sheet Range:	Datos_Tesis Baneral A1.010	* 5	Max Rows First Row Skip:	as Names a Viewer) Desca, tel :	

Fuente: elaboración propia, empleando RStudio.

Apéndice 3. Creación de resumen descriptivo por categoría

```
#CREACION DE RESUMEN ESTADISTICO
#Creacion de subset
RESUMEN <- subset(Datos_Tesis,Periodo != "Total Periodo",select = -c(ID) )
#Creando Resumen agrupado por categorias
by(RESUMEN,RESUMEN$Categoria,summary)</pre>
```

Fuente: elaboración propia, empleando RStudio.

Apéndice 4. Script de creación de EDA y Outliers por categoría

```
#CREACION DE RESUMEN ESTADISTICO
#Creacion de subset
RESUMEN <- subset(Datos_Tesis,Periodo != "Total Periodo",select = -c(ID)</pre>
)
#Creando Resumen agrupado por categorias
by(RESUMEN, RESUMEN$Categoria, summary)
#EDA categorias
install.packages("magrittr") # package installations are only needed the
first time you use it
                             # alternative installation of the %>%
install.packages("dplyr")
install.packages("tidyverse")
install.packages("ggplot2")
install.packages("ggpubr")
install.packages("rlist")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
```

```
install.packages("forcats")
install.packages("BSDA")
install.packages("Tidyverse")
library(magrittr)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(rlist)
library(purrr)
library(tibble)
library(stringr)
library(forcats)
library(BSDA)
library("car")
library(tidyverse)
#Comercio y reparacion de vehiculos
CRV <- subset(Datos_Tesis,Categoria=="Comercio y reparacion de vehiculos"</pre>
& Periodo != "Total Periodo" ,select = -c(ID))
EDA(CRV$`Valor Q(M)`)
par(mfrow=c(1,1))
boxplot(CRV$`Valor
                    Q(M)`~CRV$Año,data=CRV,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
par(mfrow=c(2,2))
hist(CRV$`Valor Q(M)`, main="Histograma de Comercio y reparacion de
vehiculos")
plot(density(CRV$`Valor Q(M)`), main="Densidad en Comercio y reparacion
de vehiculos")
boxplot(CRV$`Valor Q(M)`, main="Boxplot de Comercio y reparacion de
vehiculos")
qqPlot(CRV$`Valor Q(M)`,main="QQPlot de Comercio y reparacion
                                                                      de
vehiculos")
```

```
#Agricultura, ganaderia, silvicultura y pesca
AGSP
              subset(Datos_Tesis,Categoria=="Agricultura,
        <-
                                                             ganaderia,
silvicultura y pesca" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(AGSP$`Valor Q(M)`, main="Histograma de Agricultura, ganaderia,
silvicultura y pesca")
plot(density(AGSP$`Valor
                                    main="Densidad
                          Q(M)`),
                                                      en
                                                           Agricultura,
ganaderia, silvicultura y pesca")
boxplot(AGSP$`Valor Q(M)`, main="Boxplot de Agricultura, ganaderia,
silvicultura y pesca")
qqPlot(AGSP$`Valor
                    Q(M)`,main="QQPlot de
                                              Agricultura,
                                                             ganaderia,
silvicultura y pesca")
par(mfrow=c(1,1))
boxplot(AGSP$`Valor Q(M)`~AGSP$Año,data=AGSP,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
#Explotación de minas y canteras
EMC <- subset(Datos_Tesis,Categoria=="Explotación de minas y canteras" &
Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(EMC^{\circ} Valor Q(M), main="Histograma de Explotación de minas y
canteras")
plot(density(EMC$`Valor Q(M)`), main="Densidad en Explotación de minas y
canteras")
boxplot(EMC$`Valor Q(M)`, main="Boxplot de Explotación de minas y
canteras")
qqPlot(EMC$`Valor Q(M)`,main="QQPlot de Explotación de minas y canteras")
par(mfrow=c(1,1))
boxplot(EMC$`Valor
                    Q(M)`~EMC$Año,data=EMC,main="Ingresos
                                                            VS
                                                                 Años",
xlab="Años", ylab="Ingresos Q")
```

```
#Industrias manufactureras
IM <- subset(Datos_Tesis,Categoria=="Industrias manufactureras" & Periodo</pre>
!= "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(IM$`Valor Q(M)`, main="Histograma de Industrias manufactureras")
plot(density(IM$`Valor
                          Q(M)`),
                                    main="Densidad
                                                        en
                                                              Industrias
manufactureras")
boxplot(IM$`Valor Q(M)`, main="Boxplot de Industrias manufactureras")
qqPlot(IM$`Valor Q(M)`,main="QQPlot de Industrias manufactureras")
par(mfrow=c(1,1))
boxplot(IM$`Valor
                    Q(M)`~IM$Año,data=IM,main="Ingresos vs
                                                                  Años",
xlab="Años", ylab="Ingresos Q")
#Suministro de electricidad, agua y saneamiento
SEAS <- subset(Datos_Tesis,Categoria=="Suministro de electricidad, agua</pre>
y saneamiento" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(SEAS$`Valor Q(M)`, main="Histograma de Suministro de electricidad,
agua y saneamiento")
plot(density(SEAS$`Valor Q(M)`),
                                    main="Densidad en Suministro de
electricidad, agua y saneamiento")
boxplot(SEAS$`Valor Q(M)`, main="Boxplot de Suministro de electricidad,
agua y saneamiento")
qqPlot(SEAS$`Valor Q(M)`,main="QQPlot de Suministro de electricidad, agua
y saneamiento")
par(mfrow=c(1,1))
boxplot(SEAS$`Valor Q(M)`~SEAS$Año,data=SEAS,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
```

```
#Construcción
CON <- subset(Datos Tesis, Categoria=="Construcción" & Periodo != "Total
Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(CON$`Valor Q(M)`, main="Histograma de Construcción")
plot(density(CON$`Valor Q(M)`), main="Densidad en Construcción")
boxplot(CON$`Valor Q(M)`, main="Boxplot de Construcción")
qqPlot(CON$`Valor Q(M)`,main="QQPlot de Construcción")
par(mfrow=c(1,1))
boxplot(CON$`Valor
                    Q(M)`~CON$Año,data=CON,main="Ingresos
                                                             VS
                                                                  Años",
xlab="Años", ylab="Ingresos Q")
#Transporte y almacenamiento
TA <- subset(Datos_Tesis,Categoria=="Transporte y almacenamiento" &</pre>
Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(TA$`Valor Q(M)`, main="Histograma de Transporte y almacenamiento")
plot(density(TA$`Valor
                         Q(M)`), main="Densidad
                                                    en
                                                          Transporte
                                                                       V
almacenamiento")
boxplot(TA$`Valor Q(M)`, main="Boxplot de Transporte y almacenamiento")
qqPlot(TA$`Valor Q(M)`,main="QQPlot de Transporte y almacenamiento")
par(mfrow=c(1,1))
boxplot(TA$`Valor
                     Q(M)`~TA$Año,data=TA,main="Ingresos
                                                            VS
                                                                  Años",
xlab="Años", ylab="Ingresos Q")
#Actividades de alojamiento y servicio de comidas
AASC <- subset(Datos Tesis, Categoria=="Actividades de alojamiento y de
servicio de comidas" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(AASC$`Valor Q(M)`, main="Histograma de Actividades de alojamiento y
servicio de comidas")
```
```
plot(density(AASC$`Valor Q(M)`),
                                   main="Densidad en Actividades
                                                                     de
alojamiento y servicio de comidas")
boxplot(AASC$`Valor Q(M)`, main="Boxplot de Actividades de alojamiento y
servicio de comidas")
qqPlot(AASC$`Valor Q(M)`,main="QQPlot de Actividades de alojamiento y
servicio de comidas")
par(mfrow=c(1,1))
boxplot(AASC$`Valor Q(M)`~AASC$Año,data=AASC,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
#Información y comunicaciones
IC <- subset(Datos Tesis,Categoria=="Información y comunicaciones" &</pre>
Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(IC$`Valor Q(M)`, main="Histograma de Información y comunicaciones")
plot(density(IC$`Valor
                        Q(M)`), main="Densidad
                                                   en
                                                        Información
                                                                      У
comunicaciones")
boxplot(IC$`Valor Q(M)`, main="Boxplot de Información y comunicaciones")
qqPlot(IC$`Valor Q(M)`,main="QQPlot de Información y comunicaciones")
par(mfrow=c(1,1))
boxplot(IC$`Valor
                    Q(M)`~IC$Año,data=IC,main="Ingresos vs
                                                                 Años",
xlab="Años", ylab="Ingresos Q")
#Actividades financieras y de seguros
AFS <- subset(Datos_Tesis,Categoria=="Actividades financieras y de
seguros" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(AFS$`Valor Q(M)`, main="Histograma de Actividades financieras y de
seguros")
plot(density(AFS$`Valor Q(M)`), main="Densidad en Actividades financieras
y de seguros")
```

```
boxplot(AFS$`Valor Q(M)`, main="Boxplot de Actividades financieras y de
seguros")
qqPlot(AFS$`Valor Q(M)`,main="QQPlot de Actividades financieras y de
seguros")
par(mfrow=c(1,1))
boxplot(AFS$`Valor Q(M)`~AFS$Año,data=AFS,main="Ingresos vs
                                                                Años",
xlab="Años", ylab="Ingresos Q")
#Actividades inmobiliarias
AI <- subset(Datos_Tesis,Categoria=="Actividades inmobiliarias" & Periodo
!= "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(AI$`Valor Q(M)`, main="Histograma de Actividades inmobiliarias")
                         Q(M)`), main="Densidad
plot(density(AI$`Valor
                                                     en
                                                           Actividades
inmobiliarias")
boxplot(AI$`Valor Q(M)`, main="Boxplot de Actividades inmobiliarias")
qqPlot(AI$`Valor Q(M)`,main="OOPlot de Actividades inmobiliarias")
par(mfrow=c(1,1))
boxplot(AI$`Valor
                    Q(M)`~AI$Año,data=AI,main="Ingresos
                                                          VS
                                                                Años",
xlab="Años", ylab="Ingresos Q")
#Actividades profesionales científicas y técnicas
APCT
       <-
             subset(Datos_Tesis,Categoria=="Actividades
                                                        profesionales
científicas y técnicas" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(APCT$`Valor Q(M)`, main="Histograma de Actividades profesionales
científicas y técnicas")
plot(density(APCT$`Valor
                         Q(M)`),
                                     main="Densidad
                                                      en
                                                           Actividades
profesionales científicas y técnicas")
boxplot(APCT$`Valor Q(M)`, main="Boxplot de Actividades profesionales
científicas y técnicas")
qqPlot(APCT$`Valor Q(M)`,main="QQPlot de Actividades profesionales
científicas y técnicas")
```

```
par(mfrow=c(1,1))
boxplot(APCT$`Valor Q(M)`~APCT$Año,data=APCT,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
#Actividades de servicios administrativos y de apoyo
ASAA
            subset(Datos Tesis,Categoria=="Actividades de
       < -
                                                             servicios
administrativos y de apoyo" & Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(ASAA$`Valor Q(M)`, main="Histograma de Actividades de servicios
administrativos y de apoyo")
plot(density(ASAA$`Valor Q(M)`), main="Densidad en Actividades de
servicios administrativos y de apoyo")
boxplot(ASAA$`Valor Q(M)`, main="Boxplot de Actividades de servicios
administrativos y de apoyo")
qqPlot(ASAA$`Valor Q(M)`,main="QQPlot de Actividades de servicios
administrativos y de apoyo")
par(mfrow=c(1,1))
boxplot(ASAA$`Valor Q(M)`~ASAA$Año,data=ASAA,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
#Administración pública y defensa
APF <- subset(Datos_Tesis,Categoria=="Administración pública y defensa"
& Periodo != "Total Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(APF$`Valor Q(M)`, main="Histograma de Administración pública y
defensa")
plot(density(APF$`Valor Q(M)`), main="Densidad en Administración pública
y defensa")
boxplot(APF$`Valor Q(M)`, main="Boxplot de Administración pública y
defensa")
qqPlot(APF$`Valor Q(M)`,main="QQPlot de Administración
                                                           pública
                                                                    У
defensa")
```

```
par(mfrow=c(1,1))
                    Q(M)`~APF$Año,data=APF,main="Ingresos vs
boxplot(APF$`Valor
                                                                 Años",
xlab="Años", ylab="Ingresos Q")
#Enseñanza
EN <- subset(Datos Tesis, Categoria=="Enseñanza" & Periodo != "Total
Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(EN$`Valor Q(M)`, main="Histograma de Enseñanza")
plot(density(EN$`Valor Q(M)`), main="Densidad en Enseñanza")
boxplot(EN$`Valor Q(M)`, main="Boxplot de Enseñanza")
qqPlot(EN$`Valor Q(M)`,main="QQPlot de Enseñanza")
par(mfrow=c(1,1))
boxplot(EN$`Valor
                    Q(M)`~EN$Año,data=EN,main="Ingresos vs
                                                                 Años",
xlab="Años", ylab="Ingresos Q")
#Salud
Salud <- subset(Datos_Tesis,Categoria=="Salud" & Periodo != "Total
Periodo" ,select = -c(ID))
par(mfrow=c(2,2))
hist(Salud$`Valor Q(M)`, main="Histograma de Salud")
plot(density(Salud$`Valor Q(M)`), main="Densidad en Salud")
boxplot(Salud$`Valor Q(M)`, main="Boxplot de Salud")
qqPlot(Salud$`Valor Q(M)`,main="QQPlot de Salud")
par(mfrow=c(1,1))
boxplot(Salud$`Valor Q(M)`~Salud$Año,data=Salud,main="Ingresos vs Años",
xlab="Años", ylab="Ingresos Q")
```

Fuente: elaboración propia, empleando RStudio.

ANEXO

Anexo 1. Instalacion de Forecasting ARIMA en Microsoft Power Bi

Power BI Visuals

MARKETPLACE | MY ORGANIZATION

Add-ins may access personal and document information. By using an add-in, you agree to its Permissions, License Terms and Privacy Policy.

forecasting $ ho$		Suggested for you 🗸	
Category All		Time Series Forecasting Chart Using exponential smoothing model to predict future values based on previously observed values	Add
Advanced Analytics		★☆☆☆☆	
Data Visualizations			
Editor's Picks		Forecasting with ARIMA Predict future values based on historical data using	Add
Filters	m	Autoregressive Integrated Moving Avg (ARIMA)	
Gauges	-	****	
Infographics			
KPIs		Forecasting TBATS Time-series forecasting for series that exhibit multiple seasonalities using the TBATS model.	Add
Maps	~~~		
Time			
	MA	Forecast using Neural Network by MAQ So Use the Neural Network algorithm to forecast future values based on historical data.	Add

Fuente: Breaking-BI. Data Science in Microsoft Power Bi: Forecasting. https://2.bp.blogspot.com/-FKtyoKc_KUA/WxqguXdmM8I/AAAAAAA FmA/eG8Xzb5KffMAQabcTL0oIG_aVA2KInipACLcBGAs/s640/Forecasting%2BVisuals.png. Consulta: 23 de abril de 2022.