

Café de Colombia, análisis de la producción y la exportación en los últimos veinte años

Andrés Felipe Jaramillo Lozano
Carlos Andrés Archila Padilla
Carlos Mario Latorre Martínez
Christian David Home Acero
Darío Alejandro Durán Barrera
Felipe Herrera Cuellar
Guillermo Stiven Murillo Pantoja
Jonatan Samuel Castro Castro
Juan Esteban Álvarez David
Juan Sebastián González Quintero
Julieth Fabiana Giraldo Suarez
Mario Perlaza Erazo
Rodrigo Valdés Capote
Walter Cardona

Ficha 2672233

Servicio Nacional de Aprendizaje - SENA

Regional Valle

Centro de Electricidad y Automatización Industrial - CEAI
Técnico en Programación para Analítica de Datos - TPAD 7

Instructor

Ing. Luis Armando Amaya Quiroga

22 de septiembre de 2023

Tabla de contenido

Lista de tablas y figuras	3
Resumen	5
1. Introducción	6
2. Planteamiento del problema.....	7
2.1 Identificación del problema	7
3. Justificación	8
4. Objetivo general y objetivos específicos.....	9
4.1 Objetivo general	9
4.2 Objetivos específicos.....	9
5. Alcance	10
6. Beneficiarios	11
6.1 Beneficiarios directos.....	11
6.2 Beneficiarios indirectos	11
7. Impacto social, económico, ambiental, tecnológico	12
7.1 Impacto social.....	12
7.2 Impacto económico	12
7.3 Impacto ambiental	12
7.4 Impacto tecnológico.....	12
8. Restricciones y alternativas de solución	13
8.1 Restricciones	13
8.2 Alternativas de Solución	13
9. Resultados del desarrollo.....	14
9.1 Metodología empleada	14
9.2 Requerimientos de los datos	15
9.3 Entendimientos de los datos.....	16
9.4 Preparación de los datos	21
9.5 Preparación del modelo.....	24
9.6 Evaluación del modelo.....	36
10. Entregables y su descripción.....	41
11. Conclusiones	42
Referencias bibliográficas	46

Lista de tablas y figuras

Tabla 1.	17
Ilustración 1.	14
Ilustración 2.	18
Ilustración 3.	18
Ilustración 4.	19
Ilustración 5.	19
Ilustración 6.	20
Ilustración 7.	20
Ilustración 8.	21
Ilustración 9.	22
Ilustración 10.	22
Ilustración 11.	23
Ilustración 12.	23
Ilustración 13.	24
Ilustración 14.	25
Ilustración 15.	25
Ilustración 16.	26
Ilustración 17.	26
Ilustración 18.	27
Ilustración 19.	27
Ilustración 20.	28
Ilustración 21.	28
Ilustración 22.	29
Ilustración 23.	29
Ilustración 24.	30
Ilustración 25.	30
Ilustración 26.	31
Ilustración 27.	31
Ilustración 28.	32
Ilustración 29.	32
Ilustración 30.	33
Ilustración 31.	34
Ilustración 32.	35
Ilustración 33.	35
Ilustración 34.	35
Ilustración 35.	36
Ilustración 36.	36
Ilustración 37.	37
Ilustración 38.	37
Ilustración 39.	38
Ilustración 40.	38
Ilustración 41.	38
Ilustración 42.	39

Ilustración 43.....	39
Ilustración 44.....	40
Ilustración 45.....	40
Ilustración 46.....	40
Ilustración 47.....	42
Ilustración 48.....	43

Resumen

La economía colombiana se ha beneficiado de la producción y comercialización del café, un producto agrícola arraigado a su historia y cultura. Además, el café colombiano es reconocido a nivel mundial por su calidad y variedad, lo que ha contribuido a su destacada participación en el mercado internacional. Este proyecto se propuso analizar y comprender el desarrollo de la producción y la exportación del café colombiano durante el periodo comprendido entre 2003 y 2022, enfocándonos en compilar las dinámicas que caracterizan esta industria, y los factores políticos que han influido en su comportamiento y han impactado en su mercado.

Los resultados de este análisis revelaron que las políticas gubernamentales desempeñan un papel importante en la producción y la exportación del café, también se destacó la importancia de considerar otros factores como el contexto social, las condiciones climáticas, etc., pues no hay una variable única que intervenga en el mercado. Se observó que en los últimos cuatro años la producción y la exportación del café ha disminuido sistemáticamente.

Sin embargo, en contraparte, en este mismo lapso ha habido un aumento progresivo en el valor de la exportación del café, posiblemente relacionado con cambios en la demanda global. Pero también se identificó un aumento significativo en el valor de la cosecha. Estos resultados sugieren la necesidad de un enfoque más completo y multidimensional al analizar el mercado cafetero en Colombia.

Este estudio contribuye al entendimiento de la dinámica del café colombiano en los últimos veinte años. Si bien se utilizó la variable de las políticas gubernamentales como un factor clave, se resalta la importancia de considerar otros aspectos para nutrir este aporte. Esta perspectiva analítica ofrece información valiosa para encaminar investigaciones que permitan la toma de decisiones futuras y la formulación de políticas en el sector cafetero colombiano.

1. Introducción

Colombia, reconocido a nivel mundial como uno de los principales productores de café, destaca por su café de alta calidad y sabor excepcional. Del mismo modo, la cultura y la economía del país se ven influenciadas de manera significativa por la industria cafetera, la cual desempeña un papel fundamental al generar empleo y contribuir de forma notoria a las exportaciones del país.

La producción de café en Colombia guarda una estrecha relación con diversas dinámicas históricas, contextuales, culturales y sociales, de tal manera que responde a los distintos acontecimientos que atraviesa el país, así como a los fenómenos ambientales, los procesos electorales, las decisiones políticas y las normativas, sin embargo, sin un análisis de datos pensar en estos hechos por sí solos daría lugar a la especulación.

Este informe analiza los datos sobre la producción y la exportación del café en el país, lo que permite organizarlos y estructurarlos de tal manera que se pueda identificar las variables que afectan e impactan en su dinámica en el mercado. Además, proporciona una visión general de estos factores con el propósito de establecer una base de referencia, sacar conocimiento de los datos y preparar el camino para un análisis más profundo en futuras investigaciones.

2. Planteamiento del problema

¿Cómo ha sido el comportamiento histórico de la producción y la exportación del café en los últimos cinco periodos presidenciales de Colombia?

2.1 Identificación del problema

El café es un recurso natural de gran importancia a nivel mundial. Más del 80% de su producción se destina al comercio internacional, siendo Colombia uno de los países más visibles en este mercado. En el año 2022, el sector cafetero colombiano aportó significativamente a su economía nacional generando \$14,5 billones de pesos (COP), gracias a su rol como país exportador. Además, según estimaciones cada día se toman 2000 millones de tazas de café en todo el mundo, remarcando su arraigo en la cultura global. El café no es solo una bebida o un producto, sino un sustento fundamental para el país, ya que genera empleo y aporta ingresos significativos por exportación. Sin embargo, la industria cafetera colombiana se enfrenta a diversos desafíos, como el cambio climático y la evolución sociopolítica local, que influyen en la estabilidad económica y social del país.

La disponibilidad de datos gubernamentales y el avance de la tecnología que permite extraerlos, tratarlos y difundirlos, ha abierto la puerta a la participación ciudadana y a la transparencia en la gestión gubernamental. Sin embargo, la extensa cantidad de datos disponibles en la web plantea desafíos en cuanto a su análisis y entendimiento. Para abordar esta complejidad, se hace necesario el uso de herramientas de análisis de datos. Estas permiten análisis predictivos, visualización de la información, análisis de tendencias y cruces de variables, información que en un futuro será valiosa para la toma de decisiones en la industria cafetera y otros sectores económicos.

El estudio de estos efectos se vuelve fundamental para Colombia ya que puede proporcionar información valiosa para la formulación de políticas futuras y ayudar a la industria cafetera a anticipar y adaptarse a cambios en su dinámica. Por eso es importante para la economía y el agro colombiano llevar a cabo esta investigación, porque la producción y la exportación del café viene en una caída sistemática en los últimos cuatro años y es fundamental conocer las causas para frenar las consecuencias negativas e impulsar la economía colombiana.

3. Justificación

El café es uno de los productos agrícolas más emblemáticos y representativos de Colombia. El país es reconocido mundialmente por la alta calidad de su café suave y por ser el tercer productor más grande a nivel global. La exportación del café desempeña un papel crucial en la economía colombiana, generando empleo y contribuyendo significativamente a los ingresos de miles de familias. Por eso se busca promover una mayor comprensión y conciencia sobre el café y su influencia en la economía colombiana.

Explicando la importancia de este producto, este proyecto se embarca en el propósito de analizar y comprender el desarrollo de la comercialización del café colombiano desde el año 2003 hasta el 2022, para destacar los factores clave que han movilizadado este sector, su impacto en la economía colombiana y en la sostenibilidad del sector cafetero.

Es de resaltar que Colombia experimentó cambios significativos en su liderazgo presidencial y por ende en su política agrícola y comercial en estos últimos veinte años. Por lo tanto, el propósito de este proyecto es analizar a través de datos cómo ciertas políticas pudieron haber influido en las tendencias de crecimiento o declive de variables específicas, como la producción y la exportación. Esto permitirá documentar el impacto de dichas políticas en la dinámica del mercado de este producto.

En conclusión, debido a la importancia del café colombiano y su arraigo en el país, esta investigación no solo se lleva a cabo para documentar y exponer lo encontrado, sino también para mostrar la perspectiva analítica que es posible tener con los datos que construye el país y dejar un referente inicial de información. Sin embargo, es una realidad que también el agro y la industria cafetera se enfrentan a desafíos importantes relacionados con factores como la sostenibilidad y el cambio climático. Por lo cual es pertinente indagar y seguir documentando y procesando todos los datos sobre las variables del café como lo es la producción y exportación, para construir mejores propuestas e intervenciones que permitan avanzar a una mejor toma de decisiones por parte de los dirigentes y veeduría de los ciudadanos.

4. Objetivo general y objetivos específicos

4.1 Objetivo general

Analizar el comportamiento histórico de la producción y la exportación del café, identificando la relevancia de los últimos cinco periodos presidenciales de Colombia.

4.2 Objetivos específicos

- Recopilar y organizar los datos relacionados con el café en Colombia.
- Garantizar que los datos obtenidos sean coherentes y estén listos para el proceso de limpieza de datos y análisis.
- Utilizar las herramientas de visualización para presentar los datos.
- Emplear las herramientas de análisis de datos para interpretar el impacto de las políticas gubernamentales en el sector cafetero.
- Elaborar conclusiones basadas en la información clave alcanzada. Esto incluirá los resultados obtenidos por medio del proceso de Machine Learning.

5. Alcance

La investigación se enfoca en analizar los últimos veinte años de datos sobre el café en Colombia, teniendo en cuenta la metodología de investigación CRISP-DM, que se consideró la más apropiada para llevar a cabo un proceso de indagación en los datos. Se pretende explorar cómo algunas de las políticas gubernamentales implementadas durante ese periodo, podrían haber influido en las variaciones observadas en dicho lapso.

Para garantizar la confiabilidad de nuestros hallazgos, se planea utilizar datos provenientes de fuentes confiables, como bases de datos gubernamentales oficiales y fuentes especializadas en el sector. Haciendo énfasis en los datos de precios, producción y exportaciones.

Uno de los objetivos es analizar cómo el comportamiento de los datos en el pasado ha sido afectado por las propuestas de los gobiernos de turno en Colombia. Para lograr este objetivo se han seleccionado las herramientas para realizar análisis estadísticos, tales como, OpenRefine, Power BI, Python y R. Estas herramientas nos permiten llevar a cabo la limpieza de los datos, visualización, las correlaciones, comportamientos y comparaciones de las distintas variables. Se busca comprender cómo la dinámica de los datos podría estar relacionada a las decisiones de los gobiernos que hubo en cada periodo de tiempo estudiado.

El resultado final será presentar un informe que resuma los hallazgos que han sido claves para la investigación, teniendo en cuenta la producción total de los sacos de café del país y sus exportaciones. Las conclusiones obtenidas junto con las predicciones realizadas aportarán una valiosa visión sobre el comportamiento de esta industria.

6. Beneficiarios

Colombia tiene aroma a café. A lo largo de la cordillera de los andes se encuentran las condiciones geográficas ideales para que este fruto se convierta en el estilo de vida de 12.387 familias cafeteras, familias que constituyen la base de la cultura agrícola de nuestro país y que por más de 200 años se han dedicado a esta hermosa tarea.

6.1 Beneficiarios directos

En primera instancia los beneficiarios directos somos todo el equipo de trabajo y desarrolladores del presente proyecto quien en el proceso de investigación hemos adquirido conocimiento acerca del café, su origen, cultura, comercialización y políticas económicas que rigen los precios.

Los siguientes beneficiarios son el SENA regional Valle, el equipo de instructores y sus directivos que usarán el producto del proyecto como referencia para consultar cómo las políticas gubernamentales influyen en la producción del café.

6.2 Beneficiarios indirectos

Hay muchos beneficiarios indirectos como: Las empresas productoras, porque al comprender las tendencias de exportación e importación, los productores pueden tomar decisiones informadas sobre la producción del café, lo que puede influir en sus ingresos y rentabilidad. Las empresas que se dedican al comercio de café colombiano pueden beneficiarse al identificar oportunidades de mercado y tomar decisiones estratégicas sobre la compra y venta de café. El gobierno colombiano puede utilizar los resultados del análisis para formular políticas comerciales y agrícolas que fomenten el crecimiento sostenible de la industria del café y aumenten los ingresos fiscales a través de las exportaciones. Las personas y organizaciones que invierten en la industria del café colombiano pueden tomar decisiones de inversión más informadas basadas en las tendencias históricas y actuales de exportación e importación.

7. Impacto social, económico, ambiental, tecnológico

7.1 Impacto social

Por medio de este proyecto se puede generar conciencia sobre la importancia de la industria cafetera en Colombia y cómo las decisiones gubernamentales pueden afectarla. Esto puede fomentar el interés y la educación por la política agrícola y económica del país entre la población e incluso hacer veeduría de las decisiones que toma el Estado.

7.2 Impacto económico

El impacto económico de este proyecto está relacionado con su objeto de estudio y tiene un valor a nivel de referencia, ya que aporta al entendimiento de la dinámica de movilidad de ciertos indicadores económicos relacionados a la producción nacional del producto bajo análisis y puede servir como punto de partida para otros estudios investigativos de este tipo.

7.3 Impacto ambiental

Este proyecto no tendrá un impacto ambiental, su valor sobre este ítem, aunque hable de un producto como el café que se cultiva, será en este aspecto solo de referencia, ya que en un inicio es solo analítico.

7.4 Impacto tecnológico

El impacto tecnológico de este proyecto se refleja en la aplicación de metodologías avanzadas como CRISP-DM y el uso de herramientas tecnológicas como OpenRefine, Power BI, Python y R. Estas herramientas permiten un análisis más eficiente y una comprensión profunda de los datos, lo que no solo beneficia a este proyecto, sino que también puede fomentar la adopción de tecnologías similares en otros campos, impulsando así la toma de decisiones basada en la recolección de datos.

8. Restricciones y alternativas de solución

8.1 Restricciones

Disponibilidad de los datos: No poder acceder a todos los datos que requerimos, de la Federación Nacional de Cafeteros, para el proceso de recopilación.

Calidad de los datos: Existe el riesgo de que los datos estén incompletos, repetidos o mal digitados, lo que podría afectar la precisión de los resultados del análisis.

Plazos ajustados: El proyecto tiene unos plazos ajustados, debido a que se trata de un entregable académico que hay que desarrollar en un trimestre.

Disponibilidad de los ambientes de formación: Debido a los conflictos y paros que se presentan en el SENA, se puede dar el caso de no poder ingresar a los ambientes de formación.

8.2 Alternativas de Solución

Ampliación de los datos: Aunque haya restricción de acceso a los datos, se pueden explorar otras fuentes de datos públicos, de fuentes oficiales, para complementar la información.

Limpieza de los datos: Implementar un proceso de limpieza de datos para eliminar datos incorrectos, incoherentes, duplicados o faltantes. Utilizar herramientas de software diseñadas para este propósito, como OpenRefine.

Gestión de plazos: Para mitigar el riesgo de plazos ajustados, se puede establecer un plan de proyecto claro con objetivos concisos y utilizar herramientas de gestión de proyectos para asegurarse de que el proyecto avance según lo planeado.

Trabajar virtualmente: Intentar con los compañeros que tienen computadores con buen rendimiento, avanzar en lo posible con las tareas designadas del proyecto.

9. Resultados del desarrollo

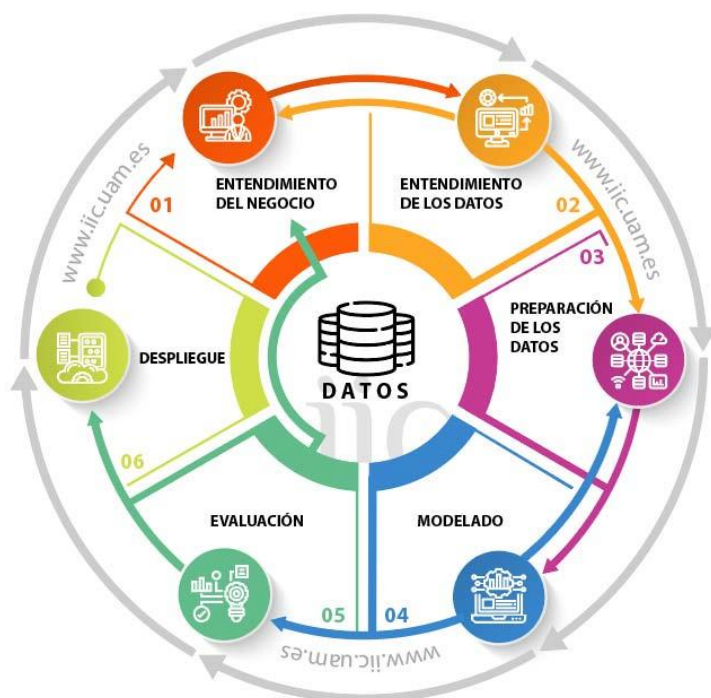
9.1 Metodología empleada

CRISP-DM (Cross-Industry Standard Process for Data Mining) es una metodología estándar utilizada para la minería de datos, un proceso que implica descubrir patrones y conocimientos útiles a partir de grandes conjuntos de datos. Fue desarrollada en 1996 por un consorcio de empresas, organizaciones y expertos en minería de datos y se ha convertido en la metodología más ampliamente utilizada para proyectos de minería de datos.

Durante más de veinte años, la metodología CRISP-DM ha sido fuente de inspiración de otros estándares como SEMMA de SAS o ASUM-DM de IBM, así como ha dado lugar a múltiples variantes que amplían o particularizan CRISP-DM a una industria o tipo de proyecto.

Ilustración 1.

CRISP-DM. Fases



- Entendimiento del negocio: Entendimiento de los objetivos y requerimientos del proyecto.
- Entendimiento de los datos: Obtención conjunto inicial de datos, exploración del conjunto de datos, identificar las características de calidad de los datos e identificar los resultados iniciales obvios. (“Metodologías aplicadas al proceso de Minería de Datos - unal.edu.co”)
- Preparación de los datos: Selección de los datos, y limpieza de los datos.
- Modelado: Implementación en herramientas de analítica de datos.
- Evaluación: Determinar si los resultados coinciden con los objetivos del negocio, e identificar los temas de negocio que deberían haberse abordado. (“Metodologías aplicadas al proceso de Minería de Datos - unal.edu.co”)
- Despliegue: Instalar los modelos resultantes en la práctica, y configuración de datos de forma repetida o continua. (“Metodologías aplicadas al proceso de Minería de Datos - unal.edu.co”)

9.2 Requerimientos de los datos

El requerimiento de datos se refiere al enfoque sistemático de reunir y medir información de diversas fuentes a fin de obtener un panorama completo y preciso de una zona de interés.

La técnica de recolección de datos utilizada para este proyecto son las fuentes abiertas: (Consiste en la información pública y gratuita que se encuentra en páginas gubernamentales, universidades, instituciones independientes, organizaciones sin fines de lucro, grandes compañías, plataformas de análisis de datos, agencias, revistas especializadas, entre otras). (“Recolección de datos: métodos, técnicas e instrumentos”)

La fuente utilizada de esta información son las estadísticas cafeteras de la Federación Nacional de Cafeteros, que contiene como base de datos los siguientes archivos de Excel: ***Precios-área-y-producción-de-café-1***, en el cual nos enfocamos en la hoja 9. Producción mensual, que contiene registros desde enero de 1956 hasta mayo del 2023 y el archivo de ***Exportaciones***, en el cual nos enfocamos en la hoja 1. Tipo volumen y hoja 2. Total valor, que contiene registros desde enero de 1958 hasta abril del 2023.

- <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>
- *Precios, área y producción del café: [Hoja 9. Producción mensual]*
- *Exportaciones de café colombiano: [Hoja 1. Tipo volumen y Hoja 2. Total valor]*

El método de recolección de datos que utilizamos para este proyecto de investigación es el método sintético: el cual es un proceso de análisis que busca resumir los aspectos más importantes de una información suministrada, que va de lo general a lo particular.

9.3 Entendimientos de los datos

Los datos seleccionados se ordenan desde el periodo más antiguo hasta el periodo más reciente, comprendiendo los últimos cinco periodos presidenciales que han finalizado. Las variables que se van a tener en cuenta serán las siguientes:

Tabla 1.**Variables**

ITEM	VARIABLE	TIPO DE VARIABLE	DESCRIPCIÓN
1	AÑO	Cuantitativa continua	Información cronológica.
2	MES	Cuantitativa continua	Información cronológica.
3	DIA	Cuantitativa continua	Información cronológica.
4	PRESIDENTE	Cualitativa	Nombre del presidente y su periodo.
5	SACOS DE CAFÉ PRODUCIDOS	Cuantitativa discreta	Representa el valor producido de café por sacos de 60 kg.
6	SACOS DE CAFÉ EXPORTADOS	Cuantitativa discreta	Representa el valor exportado de café por sacos de 60 kg.
7	VALOR EXPORTACIÓN	Cuantitativa discreta	Dato tipo moneda en dólares.
8	KILOS DE CAFÉ PRODUCIDOS	Cuantitativa discreta	Representa el valor total producido de kilos de café.
9	FACTOR DE RENDIMIENTO	Cuantitativa discreta	Porcentaje promedio de café pergamino que se necesitó para obtener café producido.
10	KILOS DE CAFÉ PERGAMINO	Cuantitativa discreta	Representa el valor de kilos de café pergamino.
11	CARGA DE CAFE	Cuantitativa discreta	Representa el valor de café pergamino por sacos de 125 kg.
12	PRECIO INTERNO DIARIO	Cuantitativa discreta	Dato tipo moneda en pesos.
13	VALOR COSECHA	Cuantitativa discreta	Dato tipo moneda en pesos.

Ilustración 2.

Precio interno diario


	Precio Interno del Café Colombiano - Diario		
	Definición: Precio interno base de compra del FoNC por carga de 125 Kg. de café pergamino seco.		
	Fuente: Grabación de Precios - Almacafé		
Fecha	Precio Interno (\$/125 Kg)	Precio Almendra Sana	Incentivo a la Calidad (\$/Kg)
vie-25-nov-22	1.963.000		
sáb-26-nov-22	1.963.000		
dom-27-nov-22	1.963.000		
lun-28-nov-22	1.930.000		
mar-29-nov-22	1.980.000		
mié-30-nov-22	1.997.000		
jue-01-dic-22	1.920.000		
vie-02-dic-22	1.895.000		
sáb-03-dic-22	1.895.000		
dom-04-dic-22	1.895.000		
lun-05-dic-22	1.920.000		
mar-06-dic-22	1.925.000		
mié-07-dic-22	1.892.000		
jue-08-dic-22	1.877.000		
vie-09-dic-22	1.865.000		

Ilustración 3.

Presidentes

N.º	Presidente		Inicio	Final
39.º		Álvaro Uribe Vélez	7 de agosto de 2002	7 de agosto de 2010
40.º		Juan Manuel Santos Calderón	7 de agosto de 2010	7 de agosto de 2018
41.º		Iván Duque Márquez	7 de agosto de 2018	7 de agosto de 2022

Ilustración 4.

Producción sacos de café


	Producción registrada - Mensual	
	Miles de sacos de 60 Kg de café verde equivalente	
	Fuente: Dirección de Investigaciones Económicas	
	Mes	Producción
	jul-19	1.317
	ago-19	1.119
	sep-19	1.088
	oct-19	1.369
	nov-19	1.506
	dic-19	1.680

Ilustración 5.

Valor de la cosecha


	Valor de la cosecha registrada -anual		
	Millones de pesos		
	Fuente: Gerencia Financiera - FNC		
Año Calendario	Valor de la cosecha	Año Cafetero	Valor de la cosecha
2000	2.279.049	2000/01	2.009.660
2001	1.959.278	2001/02	2.067.666
2002	2.120.915	2002/03	2.245.734
2003	2.244.566	2003/04	2.423.199
2004	2.668.500	2004/05	3.467.000
2005	3.457.525	2005/06	3.518.034
2006	3.606.896	2006/07	3.604.465
2007	3.818.514	2007/08	4.056.617
2008	3.825.079	2008/09	3.566.694
2009	3.400.159	2009/10	3.719.387

Ilustración 6.**Exportación sacos de café**

 Volver	Volumen de las exportaciones colombianas de café - mensual	
	Miles de sacos de 60 Kg de café verde equivalente	
	Información definitiva	
	Fuente: Federación Nacional de Cafeteros	
	MES	Total Exportaciones
	oct-21	987
	nov-21	1.135
	dic-21	1.167
	ene-22	1.032
	feb-22	983
	mar-22	1.121
	abr-22	848

Ilustración 7.**Valor de las exportaciones**

Valor de las exportaciones a todo destino- Total mensual	
Millones de dólares (USD)	
Información Preliminar	
Fuente: Federación Nacional de Cafeteros	
Mes	Valor Nominal*
sep-20	189
oct-20	219
nov-20	274
dic-20	286
ene-21	236
feb-21	283
mar-21	253
abr-21	241

9.4 Preparación de los datos

Es poco habitual que los datos crudos o en bruto sean utilizados directamente en un modelo analítico. Antes es necesario refinarlos para encontrar y resolver problemas como datos faltantes, valores no permitidos, duplicados, etc., para obtener datos de calidad con los cuales alimentar un modelo.

Ilustración 8.

Resultado. *UnCleanedCafe.csv*

Año	Mes	Día	Presidente	Sacos de cafe producidos	Sacos de cafe exportados	Valor exportacion (Dolares)	Kilos de cafe producidos	Factor de rendimiento	Kilos de cafe pergamino	Carga de cafe	Precio interno diario	Valor cosecha (Pesos)
3	ene	jue-02-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	277.000	5.383.951.356
3	ene	vie-03-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	280.875	5.459.268.365
3	ene	sáb-04-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	280.875	5.459.268.365
3	ene	dom-05-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	280.875	5.459.268.365
3	ene	lun-06-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	281.500	5.471.416.270
3	ene	mar-07-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	294.875	5.731.381.430
3	ene	mié-08-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	317.375	6.168.705.998
3	ene	jue-09-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	319.125	6.202.720.131
3	ene	vie-10-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	331.375	6.440.819.063
3	ene	sáb-11-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	331.375	6.440.819.063
3	ene	dom-12-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	331.375	6.440.819.063
3	ene	lun-13-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	330.875	6.431.100.740
3	ene	mar-14-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	331.750	6.448.107.806
3	ene	mié-15-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	337.625	6.562.298.110
3	ene	jue-16-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	337.500	6.559.868.529
3	ene	vie-17-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	337.750	6.564.727.691
3	ene	sáb-18-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	337.750	6.564.727.691
3	ene	dom-19-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	337.750	6.564.727.691
3	ene	lun-20-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	333.625	6.484.551.520
3	ene	mar-21-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	336.000	6.530.713.558
3	ene	mié-22-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	346.625	6.737.227.938
3	ene	jue-23-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	345.875	6.722.650.452
3	ene	vie-24-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	344.000	6.686.206.738
3	ene	sáb-25-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	344.000	6.686.206.738
3	ene	dom-26-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	344.000	6.686.206.738
3	ene	lun-27-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	334.250	6.496.699.425
3	ene	mar-28-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06%	2.429.581	19.437	336.125	6.533.143.139

Ilustración 9.

Proyecto OpenRefine

OpenRefine

Una potente herramienta para trabajar con datos desorganizados.

Versión nueva Descargue OpenRefine v3.7.5 ahora.

Nombre del proyecto: UncleanedCafe.csv Etiquetas

Crear proyecto →

Año	Mes	Día	Presidente	Sacos de café producidos	Sacos de café exportados	Valor exportación (Dólares)	Kilos de café producidos	Factor de rendimiento	Kilos de café pergamino	Carga de café	Precio interno diario	Valor cosecha (Pesos)
1	ene	03-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
2	ene	04-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
3	ene	05-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
4	ene	06-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
5	ene	07-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
6	ene	08-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
7	ene	09-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
8	ene	10-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
9	ene	11-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
10	ene	12-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
11	ene	13-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
12	ene	14-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
13	ene	15-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
14	ene	16-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
15	ene	17-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
16	ene	18-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
17	ene	19-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
18	ene	20-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
19	ene	21-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
20	ene	22-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
21	ene	23-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
22	ene	24-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
23	ene	25-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
24	ene	26-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
25	ene	27-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
26	ene	28-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356
27	ene	29-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76.06250%	2.429.581	19.437	277.000	5.383.951.356

Procesar los datos como

Codificación de caracteres UTF-8

Archivos CSV/TSV basados en separadores

Archivos de texto basados en renglones

Archivos de texto con campos de anchura fija

PC-Axis text files

Archivos JSON

Archivos MARC

Archivos JSON-LD

Archivos RDF-N3

Las columnas se encuentran separadas por

Comas (CSV)

Tabulaciones (TSV)

Personalizado

Usar carácter * para encerrar celdas que contengan separadores de columnas

Quitar espacios al inicio y final de las celdas

Ignorar caracteres especiales con \

Ignorar primera(s) línea(s) al inicio del archivo

Seleccionar primera(s) línea(s) para los nombres de las columnas

Nombres de columna (separados por comas)

Descartar primera(s) fila(s) de datos

Cargar al menos 0 fila(s) de datos

Procesar texto de celdas en números

Cargar filas en blanco

Cargar celdas en blanco como nulas

Cargar el origen del archivo

Almacenar archivo de almacenamiento

Actualizar previsualización

Disable auto preview

Ilustración 10.

Crear proyecto OpenRefine

Nombre del proyecto: UncleanedCafe.csv Etiquetas

Crear proyecto →

Idioma	Kilos de café pergamino	Carga de café	Precio interno diario	Valor cosecha (Pesos)
2.429.581	19.437	277.000	5.383.951.356	

Ilustración 11.**Editar celdas OpenRefine**

7294 filas

Mostrar como: **filas** registros Mostrar: 5 10 25 50 100 500 1000 filas

▼ Todo	▼ Año	▼ Mes	▼ Dia	▼ Presidente	▼ Sacos de cafe producidos	▼ Sacos de cafe exportados	▼ Valor exportacion (l
☆	1.	3	ene	jue-02-ene-03	Alvaro Uribe (P1)	30800	2.233.333
☆	2.	3	ene	vie-03-ene-03	Alvaro Uribe (P1)	30800	2.233.333
☆	3.	3	ene	sáb-04-ene-03	Alvaro Uribe (P1)	30800	
☆	4.	3	ene	dom-05-ene-03	Alvaro Uribe (P1)	30800	
☆	5.	3	ene	lun-06-ene-03	Alvaro Uribe (P1)	30800	
☆	6.	3	ene	mar-07-ene-03	Alvaro Uribe (P1)	30800	
☆	7.	3	ene	mié-08-ene-03	Alvaro Uribe (P1)	30800	
☆	8.	3	ene	jue-09-ene-03	Alvaro Uribe (P1)	25.167	
☆	9.	3	ene	vie-10-ene-03	Alvaro Uribe (P1)	25.167	
☆	10.	3	ene	sáb-11-ene-03	Alvaro Uribe (P1)	25.167	
☆	11.	3	ene	dom-12-ene-03	Alvaro Uribe (P1)	25.167	

Ilustración 12.**Transformación OpenRefine**

7294 filas

Mostrar como: **filas** registros Mostrar: 5 10 25 50 100 500 1000 filas

▼ Todo	▼ Año	▼ Mes	▼ Dia	▼ Presidente	▼ Sacos de cafe producidos	▼ Sacos de cafe exportados	▼ Valor exportacion (Dolares)	▼
☆	1.	3	ene	jue-02-ene-03	Alvaro Uribe (P1)	25167	2.233.333	1.84
☆	2.	3	ene	vie-03-ene-03	Alvaro Uribe (P1)	25167	2.233.333	1.84
☆	3.	3	ene	sáb-04-ene-03	Alvaro Uribe (P1)		2.233.333	1.84
☆	4.	3	ene	dom-05-ene-03	Alvaro Uribe (P1)			
☆	5.	3	ene	lun-06-ene-03	Alvaro Uribe (P1)			
☆	6.	3	ene	mar-07-ene-03	Alvaro Uribe (P1)			
☆	7.	3	ene	mié-08-ene-03	Alvaro Uribe (P1)			
☆	8.	3	ene	jue-09-ene-03	Alvaro Uribe (P1)	30800		
☆	9.	3	ene	vie-10-ene-03	Alvaro Uribe (P1)	30800		
☆	10.	3	ene	sáb-11-ene-03	Alvaro Uribe (P1)	30800		
☆	11.	3	ene	dom-12-ene-03	Alvaro Uribe (P1)	30800		
☆	12.	3	ene	lun-13-ene-03	Alvaro Uribe (P1)	30800	25167	
☆	13.	3	ene	mar-14-ene-03	Alvaro Uribe (P1)	30800	25167	
☆	14.	3	ene	mié-15-ene-03	Alvaro Uribe (P1)	30800	25167	

Ilustración 13.

Transformación. Data. VariablesV5.csv

Año	Mes	Día	Presidente	Sacos_de_cafe_producidos	Sacos_de_cafe_exportados	Valor_exportacion_(Dolares)	Kilos_de_cafe_producidos	Factor_de_rendimiento	Kilos_de_cafe_pergamino	Carga_de_cafe	Precio_interno_diario	Valor_cosecha_(Pesos)
2003	Enero	1/2/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	277000	5383951356
2003	Enero	1/3/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	280875	5459268365
2003	Enero	1/4/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	280875	5459268365
2003	Enero	1/5/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	280875	5459268365
2003	Enero	1/6/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	281500	5471416270
2003	Enero	1/7/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	294875	5731381430
2003	Enero	1/8/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	317375	6168705998
2003	Enero	1/9/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	319125	6202720131
2003	Enero	1/10/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	331375	6440819063
2003	Enero	1/11/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	331375	6440819063
2003	Enero	1/12/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	331375	6440819063
2003	Enero	1/13/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	330875	6431100740
2003	Enero	1/14/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	331750	6448107806
2003	Enero	1/15/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	337625	6562298110
2003	Enero	1/16/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	337500	6559868529
2003	Enero	1/17/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	337750	6564727691
2003	Enero	1/18/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	337750	6564727691
2003	Enero	1/19/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	337750	6564727691
2003	Enero	1/20/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	333625	6484551520
2003	Enero	1/21/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	336000	6530713558
2003	Enero	1/22/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	346625	6737227938
2003	Enero	1/23/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	345875	6722650452
2003	Enero	1/24/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	344000	6686206738
2003	Enero	1/25/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	344000	6686206738
2003	Enero	1/26/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	344000	6686206738
2003	Enero	1/27/2003	Alvaro Uribe_(P1)	30800	25167	2233333	1848000	0,760625	2429581	19437	334250	6496699425

9.5 Preparación del modelo

Con los datos listos para ser utilizados, se prepara un modelo descriptivo o predictivo que resuelva el problema planteado en forma de pregunta. Dependiendo del modelo, este proceso puede tomar varias iteraciones hasta alcanzar el resultado esperado.

Ilustración 14.

Dataframe. VariablesV5.csv

Año,Mes,Dia,Presidente,Sacos_de_cafe_producidos,Sacos_de_cafe_exportados,Valor_exportacion_(Dolares),Kilos

2003	Enero	1/2/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	277000	5383951356
2003	Enero	1/3/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	280875	5459268365
2003	Enero	1/4/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	280875	5459268365
2003	Enero	1/5/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	280875	5459268365
2003	Enero	1/6/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	281500	5471416270
2003	Enero	1/7/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	294875	5731381430
2003	Enero	1/8/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	317375	6168705998
2003	Enero	1/9/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	319125	6202720131
2003	Enero	1/10/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	331375	6440819063
2003	Enero	1/11/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	331375	6440819063
2003	Enero	1/12/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	331375	6440819063
2003	Enero	1/13/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	330875	6431100740
2003	Enero	1/14/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	331750	6448107806
2003	Enero	1/15/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	337625	6562298110
2003	Enero	1/16/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	337500	6559868529
2003	Enero	1/17/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	337750	6564727691
2003	Enero	1/18/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	337750	6564727691
2003	Enero	1/19/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	337750	6564727691
2003	Enero	1/20/2003	Alvaro_Uribe_(P1)	30800	25167	2233333	1848000	0.760625	2429581	19437	333625	6484551520

Ilustración 15.

Inicio Power BI

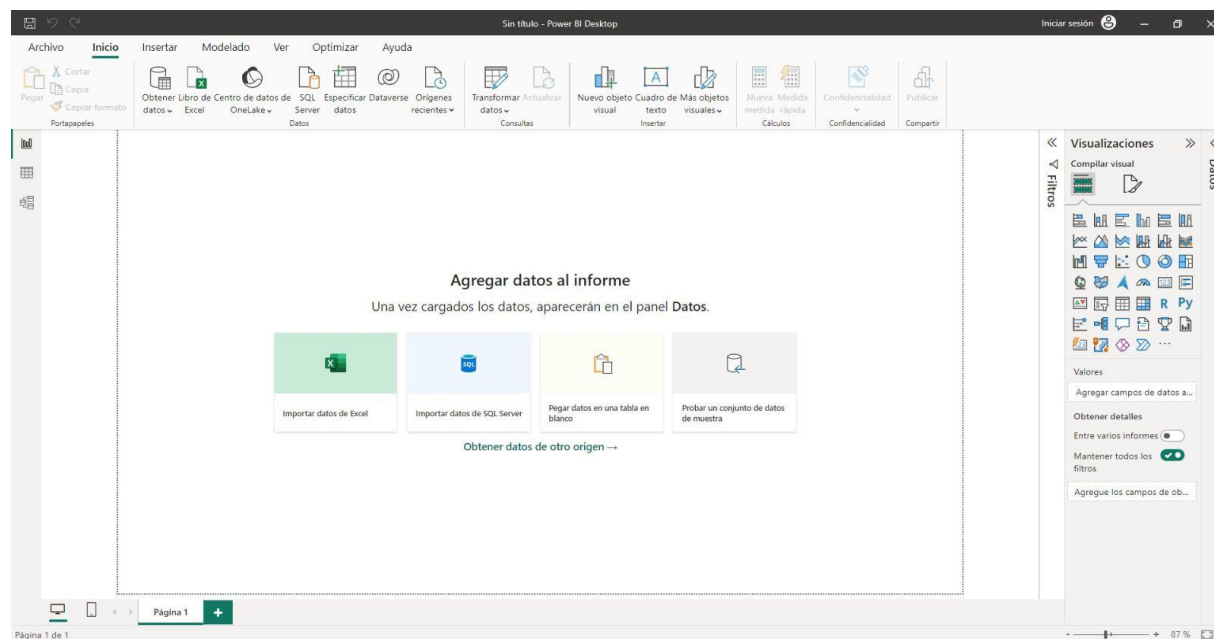


Ilustración 16.

Obtener datos Power BI

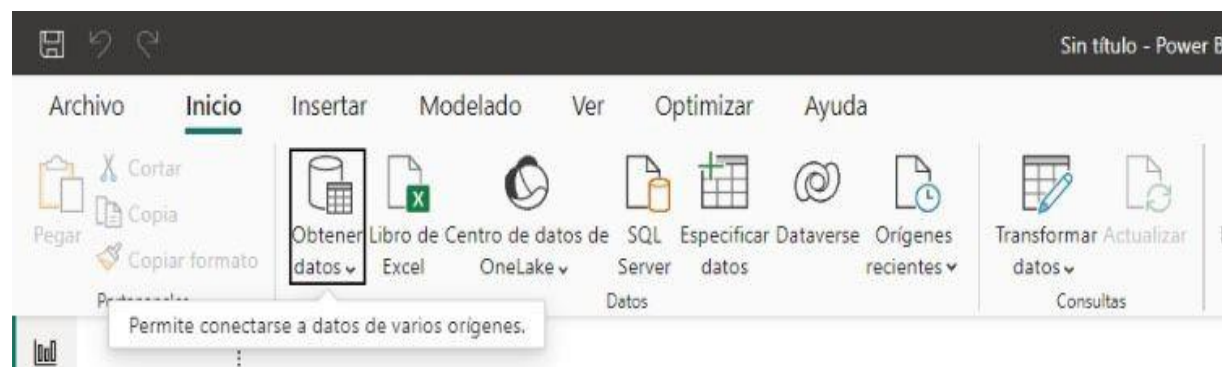


Ilustración 17.

Abrir csv Power BI

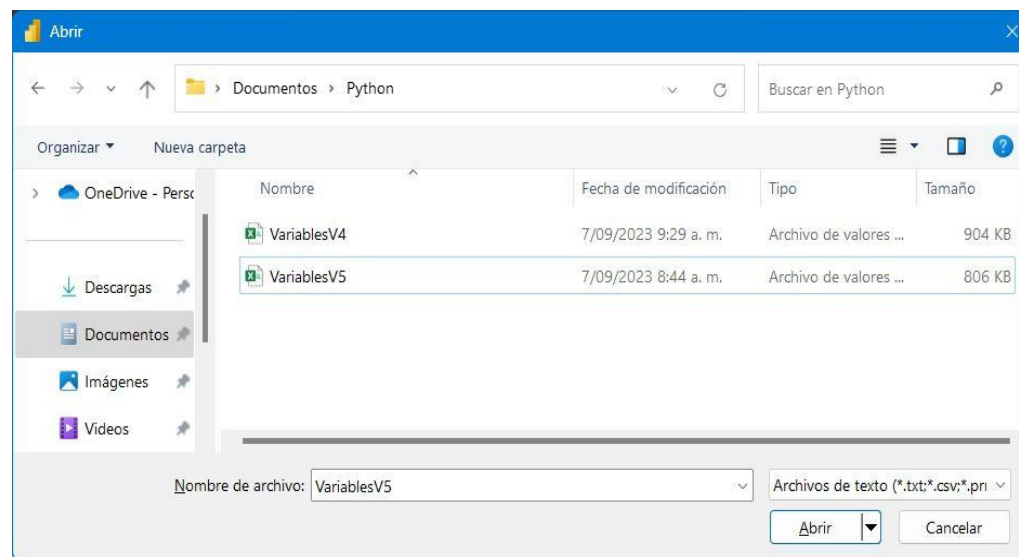


Ilustración 18.

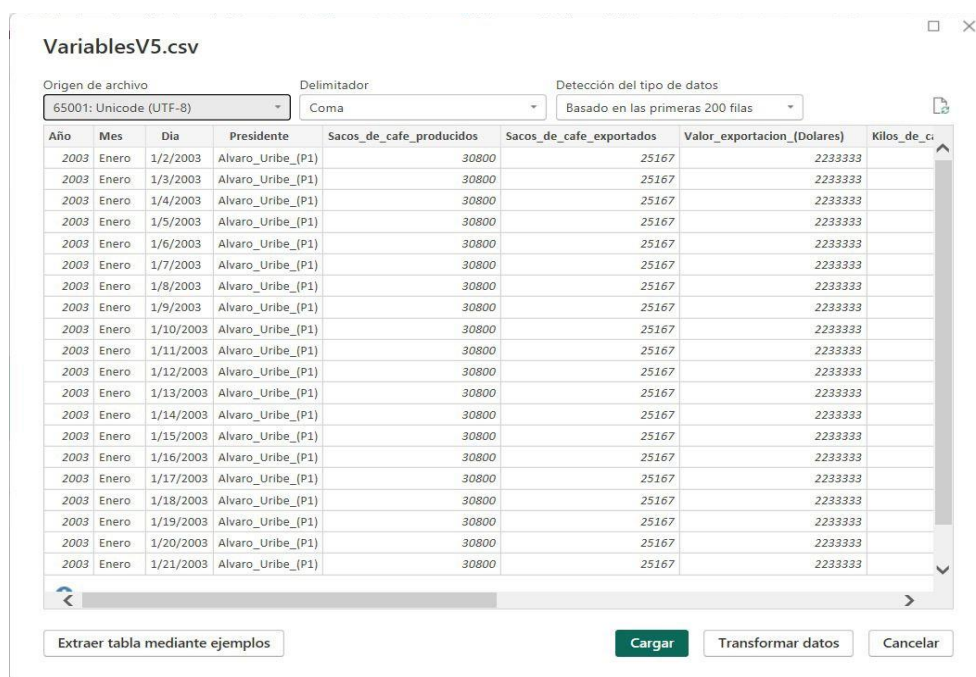
Cargar datos Power BI

Ilustración 19.

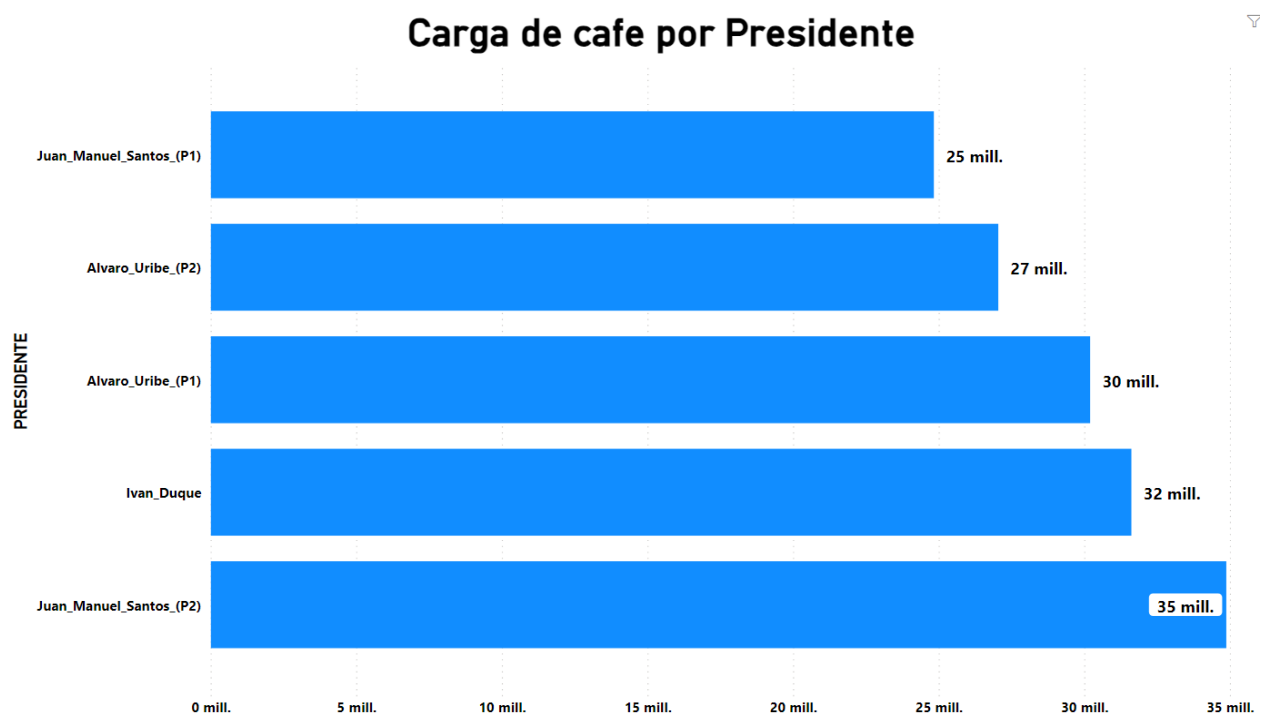
Carga de café por presidente Power BI

Ilustración 20.

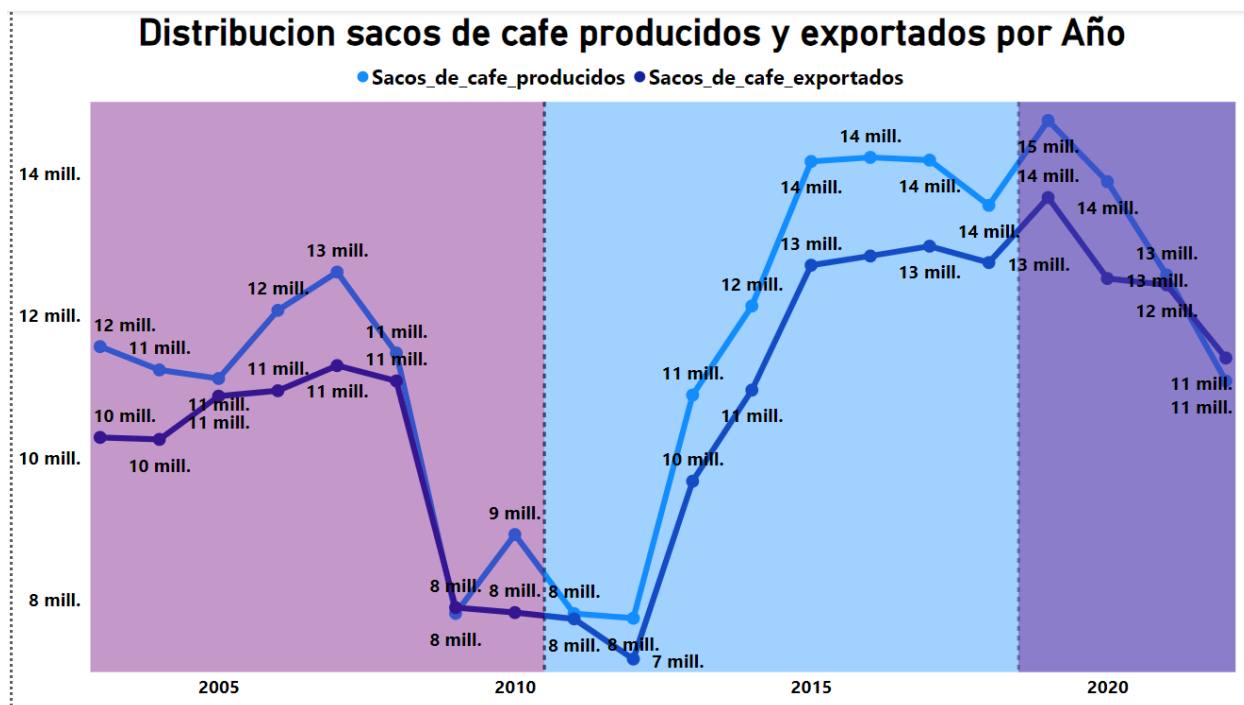
Distribución de sacos de café por año Power BI

Ilustración 21.

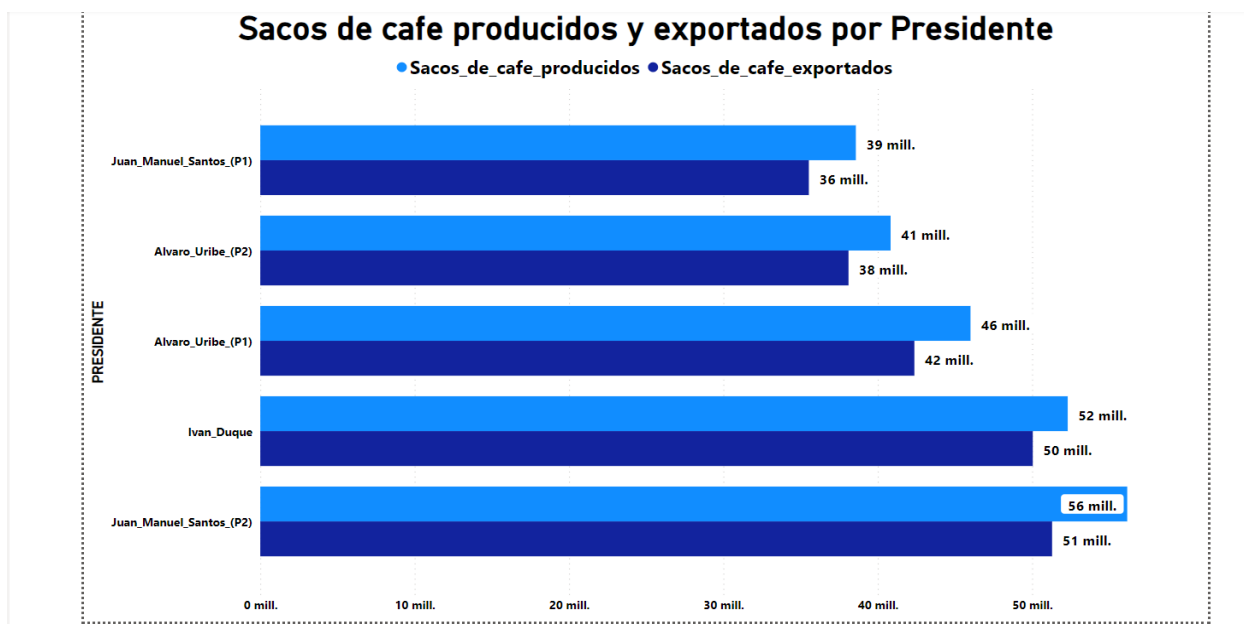
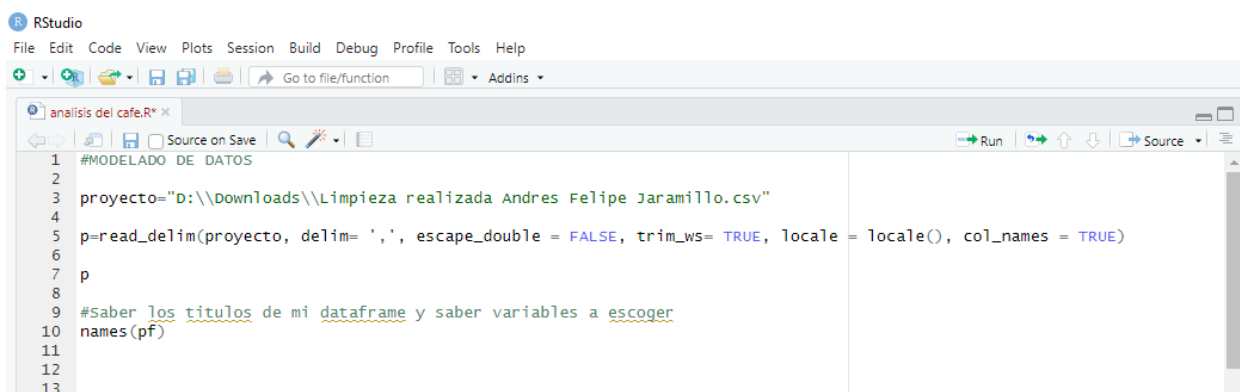
Sacos de café por presidente Power BI

Ilustración 22.

Modelado R



The screenshot shows the RStudio interface with a script editor containing the following R code:

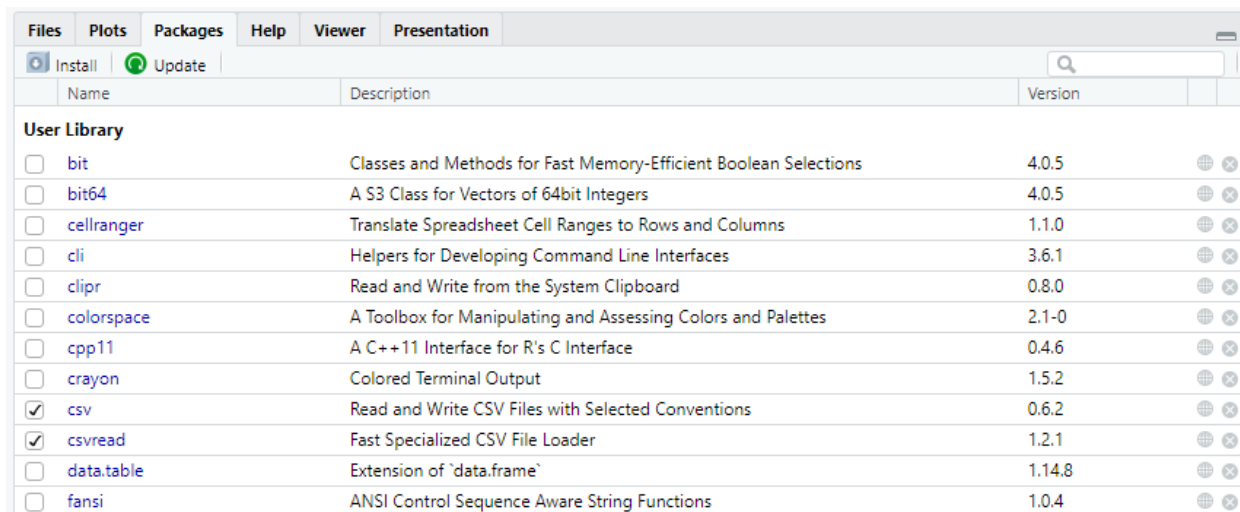
```

1 #MODELADO DE DATOS
2
3 proyecto="D:\\Downloads\\Limpieza realizada Andres Felipe Jaramillo.csv"
4
5 p=read_delim(proyecto, delim= ',', escape_double = FALSE, trim_ws= TRUE, locale = locale(), col_names = TRUE)
6
7 p
8
9 #Saber los titulos de mi dataframe y saber variables a escoger
10 names(pf)
11
12
13

```

Ilustración 23.

Librerías R



The screenshot shows the 'Packages' pane in RStudio, displaying a list of installed and available packages. The table below represents the data shown in the image.

Files		Plots	Packages	Help	Viewer	Presentation
		Install	Update			
Name	Description	Version				
User Library						
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.0.5				
<input type="checkbox"/> bit64	A S3 Class for Vectors of 64bit Integers	4.0.5				
<input type="checkbox"/> cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0				
<input type="checkbox"/> cli	Helpers for Developing Command Line Interfaces	3.6.1				
<input type="checkbox"/> clipr	Read and Write from the System Clipboard	0.8.0				
<input type="checkbox"/> colorspace	A Toolbox for Manipulating and Assessing Colors and Palettes	2.1-0				
<input type="checkbox"/> cpp11	A C++11 Interface for R's C Interface	0.4.6				
<input type="checkbox"/> crayon	Colored Terminal Output	1.5.2				
<input checked="" type="checkbox"/> csv	Read and Write CSV Files with Selected Conventions	0.6.2				
<input checked="" type="checkbox"/> csvread	Fast Specialized CSV File Loader	1.2.1				
<input type="checkbox"/> data.table	Extension of 'data.frame'	1.14.8				
<input type="checkbox"/> fansi	ANSI Control Sequence Aware String Functions	1.0.4				

Ilustración 24.

Primer Código R

```

26 #Realizar nuevo dataframe con las variables que se van a trabajar
27
28 pf=subset(p, select=c(Dia,Presidente,Sacos.de.cafe.producidos,Sacos.de.cafe.exportados,valor.exportacion..dolares.,Prec
29 pf
30
31 #Exploracion observamos que quien recibio con la produccion mas baja
32 #y aumento la produccion de cafe fue Juan Manuel Santos
33
34 fig1_sacos_producidos=plot_ly(pf, x=~Presidente, y=~Sacos.de.cafe.producidos,mode = 'lines')
35 fig1_sacos_producidos= fig1_sacos_producidos %>% layout(title = "Producción de cafe por presidente")
36 fig1_sacos_producidos
37
38 #Produccion de cafe por año
39
40 Lineaproduccion=plot_ly(pf,x=~Dia, y=~Sacos.de.cafe.producidos,name = 'trace 0', type = 'scatter', mode = 'lines')
41 Lineaproduccion= Lineaproduccion %>% layout(title = "Producción de cafe por Año")
42 Lineaproduccion
43
44 #Grafico de barra presidente sacos de cafe exportados
45
46 barraexportado=plot_ly(pf, x=~Presidente, y=~Sacos.de.cafe.exportados,mode = 'lines')
47 barraexportado= barraexportado %>% layout(title = "Exportación de cafe por presidente")
48 barraexportado
49
50 #Grafico de tiempo linea de cafe exportado
51 lineaeexportacion=plot_ly(pf,x=~Dia, y=~Sacos.de.cafe.exportados,name = 'trace 0', type = 'scatter', mode = 'lines')
52 lineaeexportacion= lineaeexportacion %>% layout(title = "Exportación de cafe por Año")
53 lineaeexportacion

```

Ilustración 25.

Producción de café por presidente R

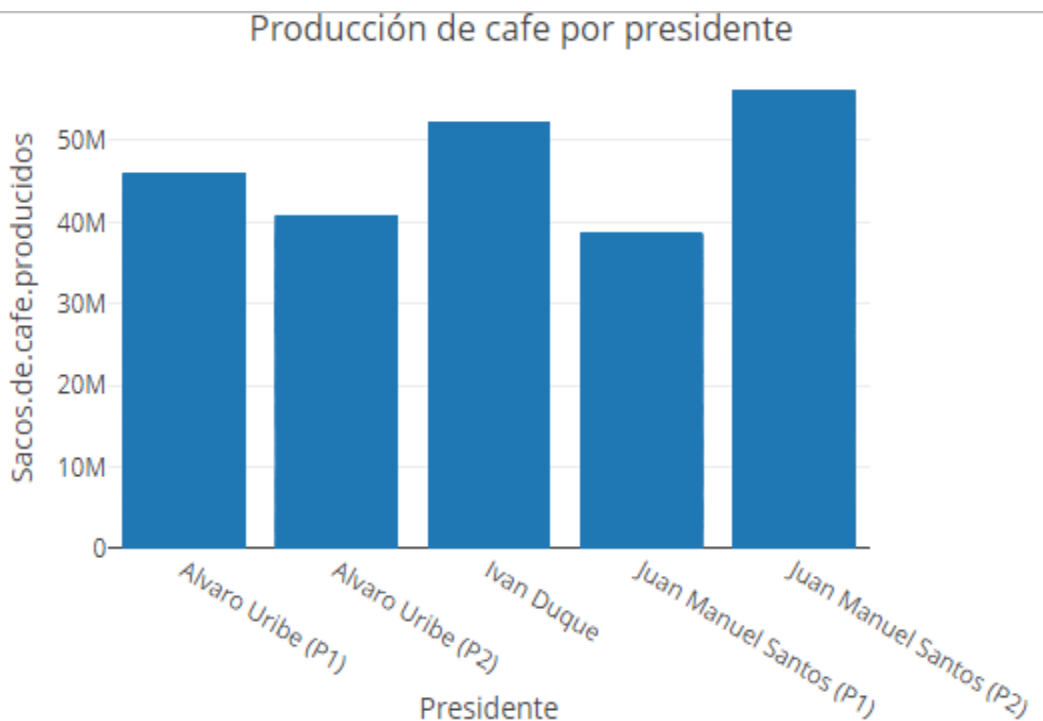
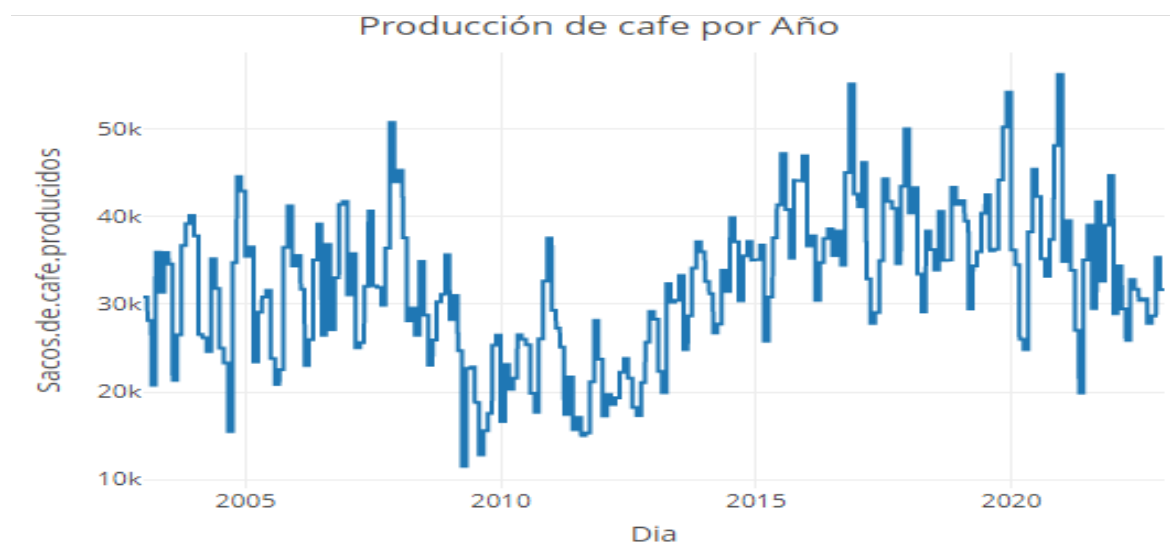


Ilustración 26.**Producción de café por año R****Ilustración 27.****Segundo Código R**

```

54
55 #Línea de tiempo con valor de valor exportacion y valor de cosecha
56
57 lineaventaycosecha=plot_ly(pf, x=~Dia, y=~Valor.exportacion..Dolares.,name = 'valor Exportacion', type = 'scatter', mod
58 lineaventaycosecha= lineaventaycosecha%% add_trace(y = ~Precio.interno.diario, name = 'valor cosecha', mode = 'lines+m
59 lineaventaycosecha= lineaventaycosecha %% layout(title = "venta en dolares y valor cosecha en dolares")
60 lineaventaycosecha
61
62 #Análisis de tiempo valor cosecha
63 pcosecha=plot_ly(pf,x=~Dia, y=~Precio.interno.diario,name = 'Precio interno', type = 'scatter', mode = 'lines')
64 pcosecha= pcosecha %% layout(title = "Precio de cosecha interno")
65 pcosecha
66

```

Ilustración 28.

Venta y valor cosecha en dólares R

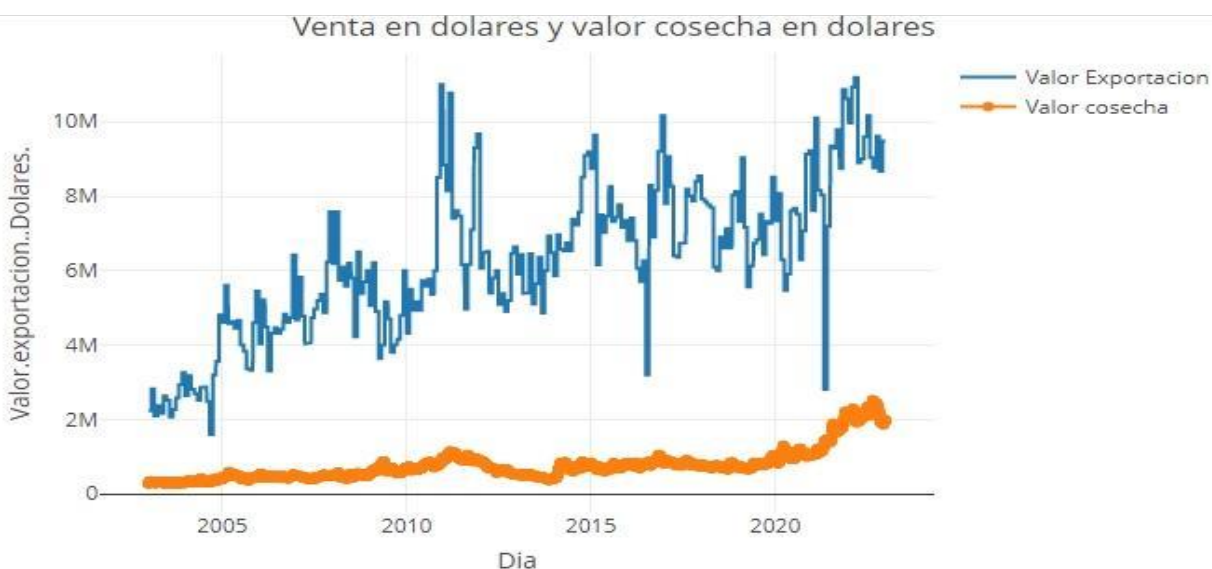


Ilustración 29.

Modelado Python

Extracion y visualizacion del dataframe se utilizara las librerias necesarias para el proceso

```
[ ] # Se cargaran las librerias necesarias para la visualizacion y extraccion de datos en el Dataframe (pandas)
# y para realizar operaciones en una matriz (Numpy)
import pandas as pd
import numpy as np
# se extrae el Dataframe y la guardamos en una variable (df). Como el dataframe tiene estacios en los encabezados
#crearemos nuestros encabezados para evitar errores de compilacion.
df=pd.read_csv('UncleanedCafe.csv')
df
```

	Año	Mes	Dia	Presidente	Sacos de cafe producidos	Sacos de cafe exportados	Valor exportacion (Dolares)	Kilos de cafe producidos	Factor de rendimiento	Kilos de cafe pergamino	Carga de cafe	Precio interno diario	Valor cosecha (Pesos)
0	3	ene	jue-02-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06250%	2.429.581	19.437	277.000	5.383.951.356
1	3	ene	vie-03-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06250%	2.429.581	19.437	280.875	5.459.268.365
2	3	ene	sáb-04-ene-03	Alvaro Uribe (P1)	30.800	25.167	2.233.333	1.848.000	76,06250%	2.429.581	19.437	280.875	5.459.268.365

dom-...

Ilustración 30.**Variables Python**

Visualización del tipo de variable de cada columna

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7294 entries, 0 to 7293
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   año                                   7294 non-null   int64
1   mes                                   7294 non-null   object
2   día                                   7294 non-null   object
3   presidente                           7294 non-null   object
4   sacos_de_cafe_producidos             7294 non-null   float64
5   sacos_de_cafe_exportados            7294 non-null   float64
6   Valor_exportacion_(Dolares)         7294 non-null   object
7   kilos_cafe_producidos               7294 non-null   object
8   x9                                    7294 non-null   object
9   kilos_cafe_pergamino                7294 non-null   object
10  carga_cafe                          7294 non-null   float64
11  precio_interno_diario               7294 non-null   object
12  valor_cosecha(pesos)                7294 non-null   object
dtypes: float64(3), int64(1), object(9)
memory usage: 740.9+ KB
```

Ilustración 31.**Transformación variables Python**

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7294 entries, 0 to 7293
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   año                                   7294 non-null   int64
1   mes                                   7294 non-null   object
2   día                                   7294 non-null   object
3   presidente                           7294 non-null   int64
4   sacos_de_cafe_producidos             7294 non-null   int64
5   sacos_de_cafe_exportados             7294 non-null   int64
6   Valor_exportacion_(Dolares)          7294 non-null   int64
7   kilos_cafe_producidos                7294 non-null   int64
8   kilos_cafe_pergamino                 7294 non-null   int64
9   carga_cafe                           7294 non-null   int64
10  precio_interno_diario                 7294 non-null   int64
11  valor_cosecha(pesos)                  7294 non-null   int64
12  factor_de_rendimiento                 7294 non-null   float64
dtypes: float64(1), int64(10), object(2)
```

Ilustración 32.

Transformación de variables Machine Learning Python

Ya con los signos y puntos eliminados podemos pasar de String a Entero para que nuestras variables sean numericas para realizar el machine learning

```
[ ] df['sacos_de_cafe_producidos'] = df['sacos_de_cafe_producidos'].astype(int)
    df['sacos_de_cafe_exportados'] = df['sacos_de_cafe_exportados'].astype(int)
    df['Valor_exportacion_(Dolares)'] = df['Valor_exportacion_(Dolares)'].astype(int)
    df['kilos_cafe_producidos'] = df['kilos_cafe_producidos'].astype(int)
    df['x9'] = df['x9'].astype(float)
    df['kilos_cafe_pergamino'] = df['kilos_cafe_pergamino'].astype(int)
    df['carga_cafe'] = df['carga_cafe'].astype(int)
    df['precio_interno_diario'] = df['precio_interno_diario'].astype(int)
    df['valor_cosecha(pesos)'] = df['valor_cosecha(pesos)'].astype(int)
    df
```

Ilustración 33.

Transformación variable factor de rendimiento Python

para los porcentajes dividiremos el dato sobre 100 con el fin de indicar que el dato de la columna 'Factor de Rendimiento' que esta ubicada en la x9 sea de manera porcentual

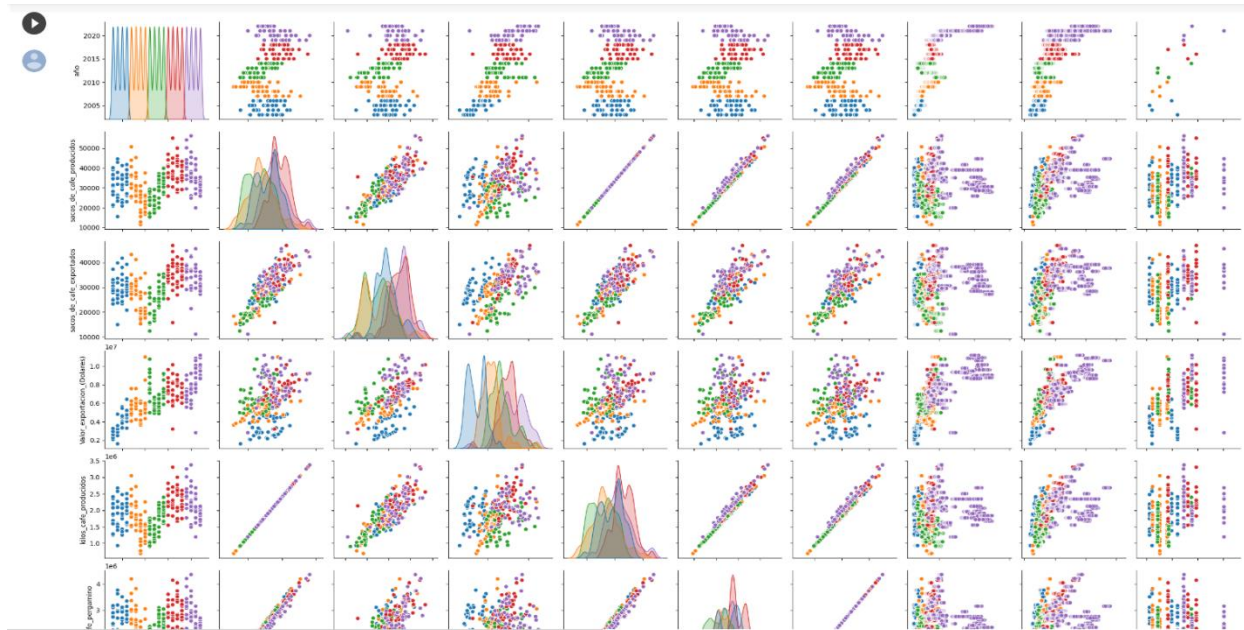
```
[ ] df['factor_de_rendimiento'] = (df['x9'] / 100)
```

Ilustración 34.

Librería seaborn Python

importamos la libreria seaborn esto con el fin de mostrar datos estadisticos con respecto a la etapa presidencial. El objetivo es visualizar que relacion tiene cada variable con la otra si esta tiene una correlacion=1 tendremos posibilidades de obtener una buena prediccion, si no hay correlacion hay mayor probabilidad de que no tenga una buena prediccion

```
import seaborn as sns
sns.pairplot(df, hue="presidente")
plt.show()
```

Ilustración 35.**Visualización de datos estadísticos Python****9.6 Evaluación del modelo****Ilustración 36.****Proceso Machine Learning**

Para realizar el proceso de ML necesitaremos repartir los datos en 90% de entrenamiento y 10 de testeo

```
datos_entrenamiento = df.sample(frac=0.9,random_state=0)
datos_test = df.drop(datos_entrenamiento.index)
```

Ilustración 37.**Datos de entrenamiento Machine Learning**

datos_entrenamiento							
	año	presidente	sacos_de_cafe_producidos	sacos_de_cafe_exportados	Valor_exportacion_(Dolares)	kilos_cafe_producidos	kilos_cafe_perg
4722	2015	4	46903	38581	7354839	2814194	36
4448	2015	4	25806	25355	6161290	1548387	19
831	2005	1	29100	28633	4633333	1746000	24
641	2004	1	34742	28645	3193548	2084516	29
2039	2008	2	23065	20710	4225806	1383871	19
...
5046	2016	4	45000	39677	8161290	2700000	33
6864	2021	5	32645	31839	8741935	1958710	22
3418	2012	3	22226	19032	5806452	1333548	18
6540	2020	5	56226	42226	9225806	3373548	43
1734	2007	2	36419	27226	4870968	2185161	30

6565 rows × 11 columns

Ilustración 38.**Datos de testeo Machine Learning**

datos_test							
	año	presidente	sacos_de_cafe_producidos	sacos_de_cafe_exportados	Valor_exportacion_(Dolares)	kilos_cafe_producidos	kilos_cafe_perg
0	2003	1	30800	25167	2233333	1848000	24
21	2003	1	30800	25167	2233333	1848000	24
24	2003	1	30800	25167	2233333	1848000	24
25	2003	1	30800	25167	2233333	1848000	24
63	2003	1	20774	23355	2096774	1246452	16
...
7250	2022	5	35333	28433	8666667	2120000	26
7252	2022	5	35333	28433	8666667	2120000	26
7261	2022	5	35333	28433	8666667	2120000	26
7268	2022	5	31645	33387	9483871	1898710	24
7272	2022	5	31645	33387	9483871	1898710	24

729 rows × 11 columns

Ilustración 39.**Librería regresión lineal Machine Learning**

importamos la librería que se usará para realizar las predicciones

```
[ ] from sklearn.linear_model import LinearRegression
    modelo = LinearRegression()
    modelo.fit(datos_entrenamiento, etiquetas_entrenamiento)
```

▼ LinearRegression
LinearRegression()

Ilustración 40.**Predicción regresión lineal Machine Learning**

Usamos la regresión lineal y adjuntamos los datos de texto para realizar las predicciones

```
predicciones = modelo.predict(datos_test)
predicciones
```

Ilustración 41.**Error porcentual Machine Learning**

importamos una librería que nos ayude a imprimir el error porcentual

```
[ ] import numpy as np
    from sklearn.metrics import mean_squared_error
    error = np.sqrt(mean_squared_error(etiquetas_test, predicciones))
    print("Error porcentual : %f" % (error*100))
```

Error porcentual : 26.473973

Ilustración 42.

Variables de predicción Machine Learning

Aquí realizamos la predicción intruduciendo informacion a las diferentes variables.

```
predic= pd.DataFrame(np.array([[23,4.0,38581.0,7354839.0,2814194.0,3603771.0,28830.0,825000.0,2.378489e+10,0.780902]]),columns=['año','presidente',
predic
```

	año	presidente	sacos_de_cafe_exportados	Valor_exportacion_(Dolares)	kilos_cafe_producidos	kilos_cafe_pergamino	carga_cafe	precio_inte
0	23.0	4.0	38581.0	7354839.0	2814194.0	3603771.0	28830.0	

```
[ ] modelo.predict(predic)

array([46927.53352226])
```

Ilustración 43.

Regresión logística Machine Learning

haremos una predicción con regresión logística para predecir el precio interno diario

```
ivan_duque=df.loc[df['presidente']==5]
ivan_duque
```

	año	presidente	sacos_de_cafe_producidos	sacos_de_cafe_exportados	Valor_exportacion_(Dolares)	kilos_cafe_producidos	kilos_cafe_per
5840	2019	5	41806	37613	7322581	2508387	3:
5841	2019	5	41806	37613	7322581	2508387	3:
5842	2019	5	41806	37613	7322581	2508387	3:
5843	2019	5	41806	37613	7322581	2508387	3:
5844	2019	5	41806	37613	7322581	2508387	3:
...
7289	2022	5	31645	33387	9483871	1898710	2:
7290	2022	5	31645	33387	9483871	1898710	2:
7291	2022	5	31645	33387	9483871	1898710	2:
7292	2022	5	31645	33387	9483871	1898710	2:
7293	2022	5	31645	33387	9483871	1898710	2:

1454 rows x 11 columns

Ilustración 44.**Regresión logística Iván Duque Machine Learning**

```
[ ] ivan_duque.to_csv('ivan.csv')
```

```
[ ]
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
from sklearn.metrics import mean_absolute_error, r2_score
```

```
▶ ivan_duque.drop(columns=['presidente'])
```

Ilustración 45.**Código regresión logística Machine Learning**

```
X = ivan_duque[['año','sacos_de_cafe_producidos','sacos_de_cafe_exportados','Valor_exportacion_(Dolares)','kilos_cafe_exportados']]
y = ivan_duque['precio_interno_diario']
```

```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.25, random_state=99)
modelo1 = LogisticRegression()
modelo1.fit(X_train,y_train)
y_predict=modelo1.predict(X_test)
y_predict_train=modelo1.predict(X_train)
mae_test = mean_absolute_error(y_test, y_predict)
r2_test = r2_score(y_test, y_predict)
mae_train = mean_absolute_error(y_train, y_predict_train)
r2_train = r2_score(y_train,y_predict_train)
print(mae_test,r2_test)
print(mae_train,r2_train)
```

Ilustración 46.**Modelo predictivo**

```
▶ modelo1.predict([[22022,28143,37613,7322581,2508387,3247579,25981,5647626247,0.760625]])
```

```
ⓘ /usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have a
  warnings.warn(
  array([692000])
```


10. Entregables y su descripción

Repositorio GitHub (publico)

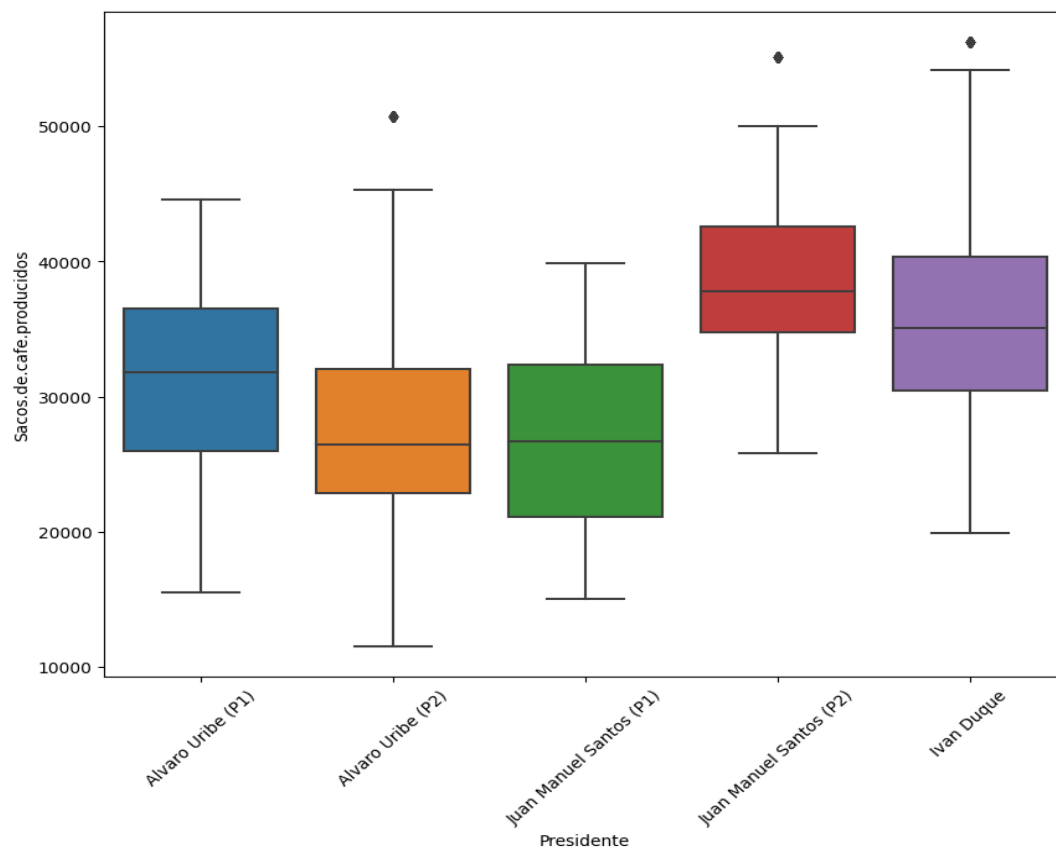
- ✓ Presentación Power Point: Presentación del proyecto
- ✓ Dashboard Power BI: Dashboard del proyecto
- ✓ Captura Dashboard JPG: Captura del dashboard
- ✓ Archivo OpenRefine: Limpieza de los datos del proyecto
- ✓ Archivo CSV: Datos del proyecto
- ✓ Graficas Power BI: Gráficas básicas del proyecto
- ✓ Archivo R: Modelado de los datos del proyecto en R
- ✓ Archivo Python Google Collab: Machine Learning de los datos del proyecto
- ✓ Informe Word: Informe del proyecto
- ✓ Documento PDF: Documento del proyecto

11. Conclusiones

En la interpretación de los sacos producidos por periodo presidencial identificamos que las medias del segundo periodo de Álvaro Uribe coinciden al comportamiento del primer periodo de Juan Manuel Santos, el primer cuartil presenta una distancia representativa y culminan casi a la par, los gráficos que presentan menor desviación son las de Juan Manuel Santos.

Ilustración 47.

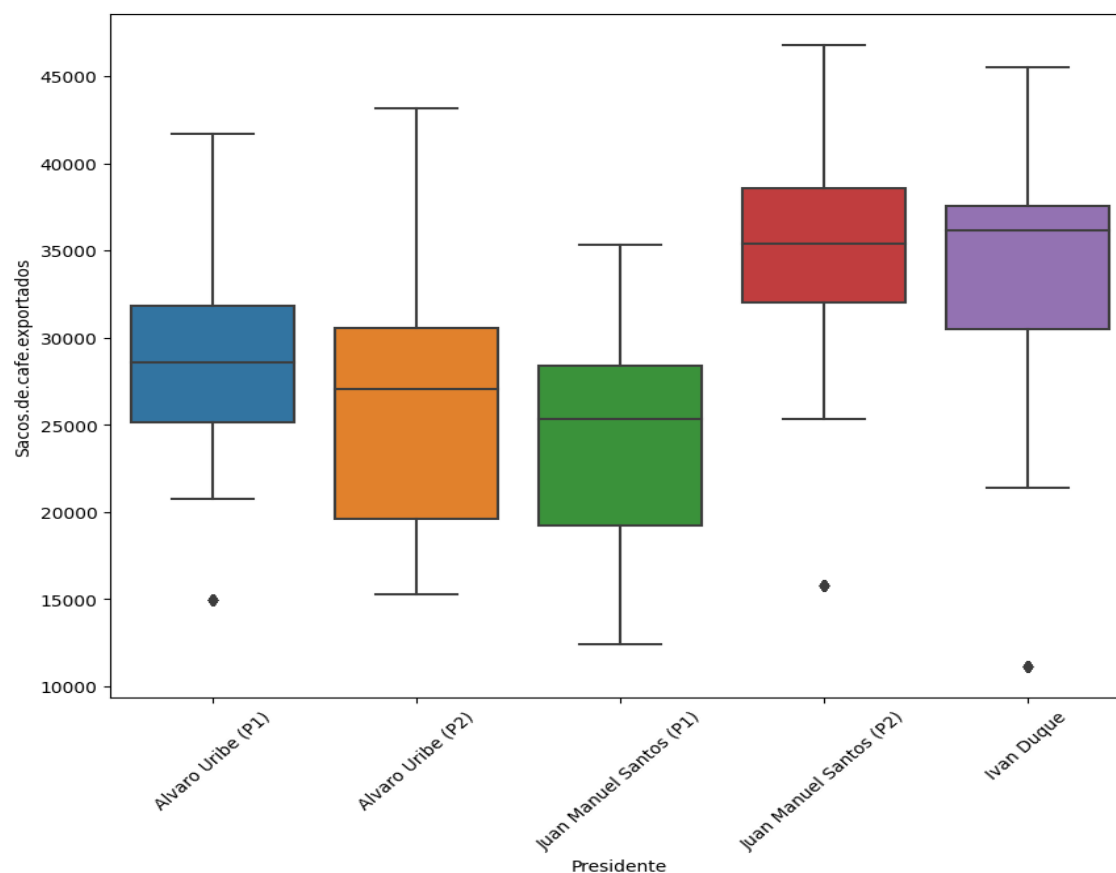
Grafica sacos producidos



Vemos en la interpretación de las exportaciones que los sacos de café se evidencia una disminución relativa por periodo presidencial desde el primer periodo de Álvaro Uribe hasta el primer periodo de Juan Manuel Santos, en el segundo periodo de Juan Manuel Santos se contempla como el cuartil 1 del primer periodo es superado por el primer cuartil del segundo periodo denotando unas diferencias significativas en exportación y en el periodo presidencial de Iván Duque fue notorio mantener la exportación con una media superior a la de Juan Manuel Santos pero que tuvo unas desviaciones más grandes.

Ilustración 48.

Grafica sacos exportados



En conclusión analizamos que en el segundo periodo presidencial de Juan Manuel Santos hubo la mayor producción y exportación de café, en contraste con su primer periodo presidencial donde fue el menor en producción y exportación de los últimos cinco periodos presidenciales.

Según indica la historia al iniciar el primer periodo de Juan Manuel Santos el panorama cafetero era muy poco alentador. Por una parte, los precios de los fertilizantes habían llegado a niveles récords, como resultado del incremento en los precios del petróleo. Esto implicó un incremento en los costos de producción.

Lo anterior, sumado a los problemas de inseguridad en muchas zonas productoras, trajo consigo el descuido e incluso el abandono de muchos cafetales y su envejecimiento. De manera adicional, se produjo un incremento considerable en el nivel de humedad debido a las altas precipitaciones que causó que los cafetales fueran altamente susceptibles a la roya.

Política gubernamental caficultura un modelo de paz: Ante los fenómenos descritos, la administración de Juan Manuel Santos se propuso lidiar con este gran reto, firmando el primer acuerdo por la prosperidad cafetera en el cual se sentaron las bases de lo que sería la acción del gobierno en este tema.

Mayor productividad: Se observó una recuperación de la producción, soportada en incrementos de productividad, como resultado de la transformación productiva y la sanidad vegetal del cultivo. Esto corresponde a unos cultivos más jóvenes, sembrados con variedades resistentes, mayor tecnificación, mejores condiciones sanitarias y sobre todo cafetales mejores preparados para enfrentar los desafíos que supone la variabilidad climática.

Con este ejemplo de implementación de política gubernamental podemos manifestar que para la industria cafetera es relevante revisar el comportamiento de la producción junto con la exportación para satisfacer la demanda del mercado, las técnicas e investigación del cultivo son prioridad para mantener un mercado competente y un producto de calidad. Los gobiernos son partícipes de la interacción del café colombiano ante el mercado internacional y cómo las políticas de gobierno afectan la estabilidad económica de los caficultores y la población que depende de la cosecha cafetera.

Es por ello por lo que podemos crear la hipótesis para determinar que, de acuerdo con los comportamientos del mercado, el comportamiento en producción y el comportamiento de exportación, las mejores políticas de gobierno para el agro cafetero fueron las del presidente Juan Manuel Santos.

Teniendo en cuenta la hipótesis planteada y los históricos de comportamiento, determinamos que aunque el trabajo realizado en los dos periodos de Juan Manuel Santos no fueron tan satisfactorios para mejorar el promedio del precio del café, se refleja un crecimiento significativo a partir de su segundo periodo para el valor promedio y que se ve reflejado en el periodo de Iván Duque, que retomando esta proyección no intensificó propuestas relevantes para aumentar la producción y la exportación que venía determinada por Juan Manuel Santos, en cambio estas producciones y exportaciones se vieron disminuidas sistemáticamente año a año, pero que se vio implicado por una crisis mundial que fue el COVID-19.

Referencias bibliográficas

Diplomado, D. (s/f). *Minería de Datos*. Edu.co. Recuperado el 22 de septiembre de 2023, de https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

Edición, C. (s/f). *La Guía del Café*. Intracen.org. Recuperado el 22 de septiembre de 2023, de https://intracen.org/sites/default/files/media/file/media_file/2022/06/29/itc_coffee_4th_report_20211029_es_web.pdf

El mercado del café en el mundo - Datos estadísticos. (s/f). Statista. Recuperado el 22 de septiembre de 2023, de <https://es.statista.com/temas/9035/el-cafe-en-el-mundo/>

Estadísticas Cafeteras. (2019, noviembre 30). Federación Nacional de Cafeteros. <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

IBM Documentation. (2021, agosto 17). lbm.com. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=objectives-e-retail-example-finding-business>

Informes archivos - Federación Nacional de Cafeteros. (s/f). Federación Nacional de Cafeteros. Recuperado el 22 de septiembre de 2023, de <https://federaciondecafeteros.org/wp/tipos/informes/>

Posada, S. G. (2019, enero 14). La economía del café: ¿Quién se está quedando el dinero? (2019). *Qué Café!* <https://quecafe.info/la-economia-del-cafe-quien-se-esta-quedando-el-dinero/>

Santos, D. (2022, septiembre 14). *Recolección de datos: métodos, técnicas e instrumentos*. Hubspot.es. <https://blog.hubspot.es/marketing/recoleccion-de-datos>

(S/f). *Federaciondecafeteros.org*. Recuperado el 22 de septiembre de 2023, de https://federaciondecafeteros.org/static/files/1La_politica_cafetera_2010-2014.pdf