

CIUDAD JUÁREZ



RETO-ASPIRANTE

DT-DS-BOSCH - Datathon

Correo:

cristian.fuente@train.ia.center

Aspirante:

Cristian Ulises DE LA FUENTE
ESPINO

29 de noviembre de 2023

Índice

llevó a cabo una observación inicial de los datos con el objetivo de abordarlos de manera efectiva. En este contexto, se optó por emplear SQLite para el manejo eficiente de los datos, simplificando así su manipulación.

Name	Type	Schema
tabla_2009	CREATE TABLE	CREATE TABLE "tabla_2009" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2010	CREATE TABLE	CREATE TABLE "tabla_2010" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2011	CREATE TABLE	CREATE TABLE "tabla_2011" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2012	CREATE TABLE	CREATE TABLE "tabla_2012" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2013	CREATE TABLE	CREATE TABLE "tabla_2013" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2014	CREATE TABLE	CREATE TABLE "tabla_2014" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2015	CREATE TABLE	CREATE TABLE "tabla_2015" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2016	CREATE TABLE	CREATE TABLE "tabla_2016" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2017	CREATE TABLE	CREATE TABLE "tabla_2017" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2018	CREATE TABLE	CREATE TABLE "tabla_2018" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2019	CREATE TABLE	CREATE TABLE "tabla_2019" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2020	CREATE TABLE	CREATE TABLE "tabla_2020" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2021	CREATE TABLE	CREATE TABLE "tabla_2021" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_2022	CREATE TABLE	CREATE TABLE "tabla_2022" ("Anio" INTEGER, "Idestado" INTEGER, "Nomestado" TEXT, "Idddr" INTEGER, "Nomddr" TEXT, "Idcac"
tabla_clean_2003	CREATE TABLE	CREATE TABLE "tabla_clean_2003" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2004	CREATE TABLE	CREATE TABLE "tabla_clean_2004" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2005	CREATE TABLE	CREATE TABLE "tabla_clean_2005" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2006	CREATE TABLE	CREATE TABLE "tabla_clean_2006" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2007	CREATE TABLE	CREATE TABLE "tabla_clean_2007" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2008	CREATE TABLE	CREATE TABLE "tabla_clean_2008" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2009	CREATE TABLE	CREATE TABLE "tabla_clean_2009" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2010	CREATE TABLE	CREATE TABLE "tabla_clean_2010" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2011	CREATE TABLE	CREATE TABLE "tabla_clean_2011" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2012	CREATE TABLE	CREATE TABLE "tabla_clean_2012" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2013	CREATE TABLE	CREATE TABLE "tabla_clean_2013" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2014	CREATE TABLE	CREATE TABLE "tabla_clean_2014" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2015	CREATE TABLE	CREATE TABLE "tabla_clean_2015" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2016	CREATE TABLE	CREATE TABLE "tabla_clean_2016" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2017	CREATE TABLE	CREATE TABLE "tabla_clean_2017" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2018	CREATE TABLE	CREATE TABLE "tabla_clean_2018" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN
tabla_clean_2019	CREATE TABLE	CREATE TABLE "tabla_clean_2019" ("Anio" INTEGER, "Idestado" INTEGER, "Idddr" INTEGER, "Icdader" INTEGER, "Idmunicipio" IN

Figura 1: Estructura de las tablas en SQLite.

Durante este proceso, se identificó que no todos los archivos CSV tenían la misma estructura. Por lo tanto, se procedió a crear nuevas tablas para limpiar los datos, con el objetivo de visualizar la información de interés de manera más clara, utilizando Python.

```
"Cierre_agr_mun_2022.csv"
}
combinaciones = [
    ('Idestado', 'Nomestado'),
    ('Idddr', 'Nomddr'),
    ('Icdader', 'Nomcader'),
    ('Idmunicipio', 'Nomunicipio'),
    ('Idciclo', 'Nomcicloproductivo'),
    ('Idmodalidad', 'Nommodalidad'),
    ('Idunidadmedida', 'Nomunidad')
]

def crear_tabla_desde_archivo(archivo, conn):
    fecha = archivo.split('.')[3].split('.')[0]
    df = pd.read_csv(f'Datos/{archivo}', encoding='ISO-8859-1')
    nombre_tabla = f'tabla_{fecha}'
    df.to_sql(nombre_tabla, conn, index=False, if_exists='replace')

def limpiar_y_crear_tabla(archivo, conn):
    fecha = archivo.split('.')[3].split('.')[0]
    df = pd.read_csv(f'Datos/{archivo}', encoding='ISO-8859-1')
    columna_precio = "Precio" if "Precio" in df.columns else "Preciomediorural"
    columnas = [
        "Anio", "Idestado", "Idddr", "Icdader", "Idmunicipio", "Idciclo",
        "Idmodalidad", "Idunidadmedida", "Idcultivo", "Sembrada", "Cosechada",
        "Siniestrada", "Volumenproduccion", "Rendimiento", columna_precio, "Valorproduccion"
    ]
    df_clean = df[columnas].rename(columns={columna_precio: "Precio"})
    tabla_clean = f'tabla_clean_{fecha}'
    df_clean.to_sql(tabla_clean, conn, index=False, if_exists='replace')

conn = sqlite3.connect('DB_AGRICOLA3.db')

for archivo in archivos:
    if f'tabla_{archivo.split('.')[3].split('.')[0]}' not in pd.read_sql_query("SELECT name FROM sqlite_master WHERE type='table';", conn)["name"].t:
        crear_tabla_desde_archivo(archivo, conn)
        limpiar_y_crear_tabla(archivo, conn)

años = range(2003, 2023)

for combinacion in combinaciones:
    nombre_columna_id, nombre_columna_nom = combinacion
    consulta_sql = f'SELECT DISTINCT {nombre_columna_id}, {nombre_columna_nom} FROM tabla_{años[0]}'

    for año in años[1:]:
        consulta_sql += f' UNION SELECT DISTINCT {nombre_columna_id}, {nombre_columna_nom} FROM tabla_{año}'

    df_resultado = pd.read_sql_query(consulta_sql, conn)
    nombre_tabla = f'{nombre_columna_id}_{nombre_columna_nom}_tabla'
    df_resultado.to_sql(nombre_tabla, conn, index=False, if_exists='replace')
    print(f'Tabla creada: {nombre_tabla}')
```

Figura 2: Limpieza de datos.

Posteriormente, se normalizaron aquellas columnas con nombres diferentes pero que representaban la misma información.

Con esta preparación, se facilitó la realización de consultas de manera organizada y eficiente. La primera consulta se centró en analizar el comportamiento anual del valor de producción de los diversos productos presentes en el mercado mexicano.

```

conn = sqlite3.connect('DB_AGRICOLA3.db')
años = range(2003, 2023)

consulta_sql3 = f'''
SELECT Idcultivo,
Anio,
SUM(Sembrada) AS SumaSembrada,
SUM(Cosechada) AS SumaCosechada,
SUM(Siniestrada) AS SumaSiniestrada,
SUM(Volumenproduccion) AS SumaVolumenproduccion,
SUM(Rendimiento) AS SumaRendimiento,
SUM(Precio) AS SumaPrecio,
SUM(Valorproduccion) AS SumaValorproduccion
FROM tabla_clean(años[0])
GROUP BY Idcultivo, Anio
'''

for año in años[1:]:
    consulta_sql3 += f'''
UNION
SELECT Idcultivo,
Anio,
SUM(Sembrada) AS SumaSembrada,
SUM(Cosechada) AS SumaCosechada,
SUM(Siniestrada) AS SumaSiniestrada,
SUM(Volumenproduccion) AS SumaVolumenproduccion,
SUM(Rendimiento) AS SumaRendimiento,
SUM(Precio) AS SumaPrecio,
SUM(Valorproduccion) AS SumaValorproduccion
FROM tabla_clean(año) WHERE Anio= {año}
GROUP BY Idcultivo
'''

print(consulta_sql3)
Ventas_por_cultivo_anuales = pd.read_sql_query(consulta_sql3, conn)
print(Ventas_por_cultivo_anuales)

conn.close()

for i in años:
    df_plot = Ventas_por_cultivo_anuales[Ventas_por_cultivo_anuales['Anio'] == i]
    df_plot_sorted = df_plot.sort_values(by='SumaValorproduccion', ascending=False)
    df_plot_top10 = df_plot_sorted.head(10)
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    sns.barplot(x='Idcultivo', y='SumaValorproduccion', data=df_plot_top10)
    plt.xlabel('Idcultivo')
    plt.ylabel('Valor de produccion')
    plt.title(f'Top 10 Productos con mayor de Valor de produccion en México durante {i}')
    plt.xticks(rotation=45)
    nombre_archivo = f"Top10_Productos_Mexico_{i}.png"
    ruta_archivo = os.path.join(output_directory, nombre_archivo)
    plt.savefig(ruta_archivo)
    plt.show()

```

Figura 3: Consulta del comportamiento anual del valor de producción.

A partir del DataFrame obtenido, se generaron gráficos básicos que representan el Top 10 de productos que generaron el mayor valor de producción en el mercado, dando como resultado las siguientes visualizaciones.

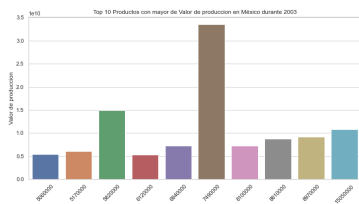


Figura 4: Top10 Productos Mexico 2003

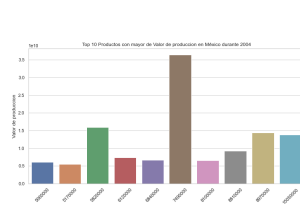


Figura 5: Top10 Productos Mexico 2004

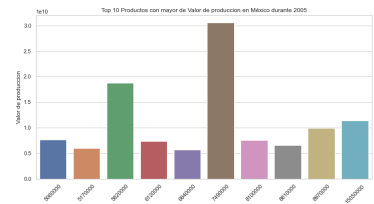


Figura 6: Top10 Productos Mexico 2005

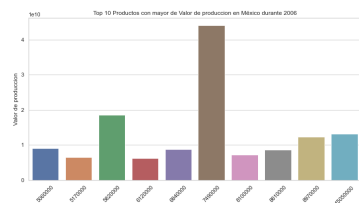


Figura 7: Top10 Productos Mexico 2006

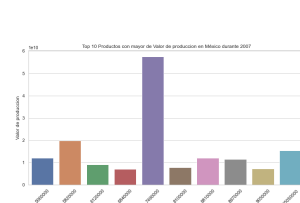


Figura 8: Top10 Productos Mexico 2007

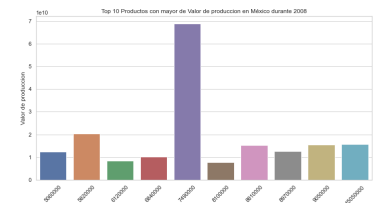


Figura 9: Top10 Productos Mexico 2008

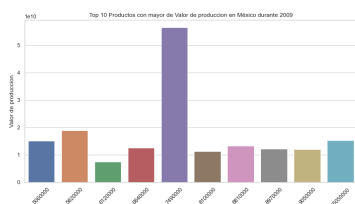


Figura 10: Top10 Productos Mexico 2009

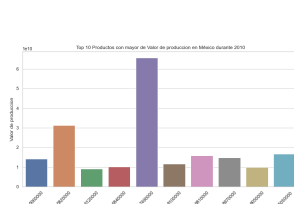


Figura 11: Top10 Productos Mexico 2010

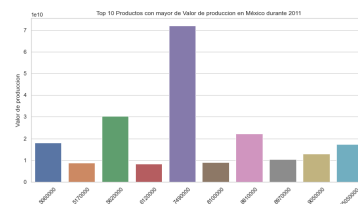


Figura 12: Top10 Productos Mexico 2011

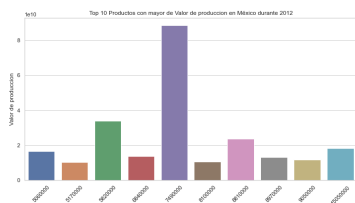


Figura 13: Top10 Productos Mexico 2012

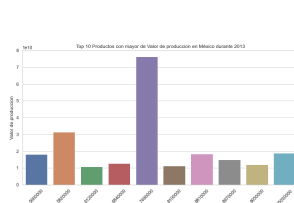


Figura 14: Top10 Productos Mexico 2013

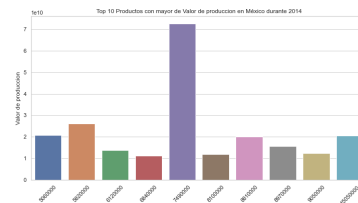


Figura 15: Top10 Productos Mexico 2014

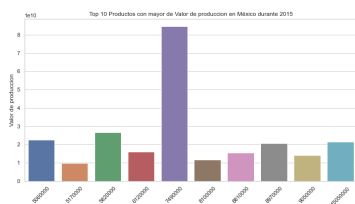


Figura 16: Top10 Productos Mexico 2015

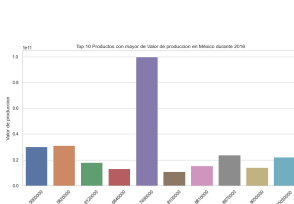


Figura 17: Top10 Productos Mexico 2016

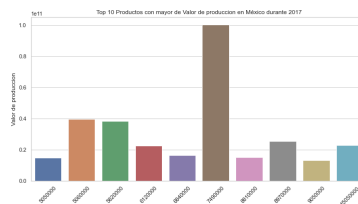


Figura 18: Top10 Productos Mexico 2017

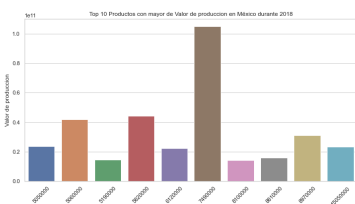


Figura 19: Top10 Productos Mexico 2018

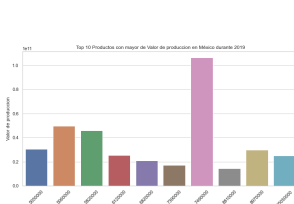


Figura 20: Top10 Productos Mexico 2019

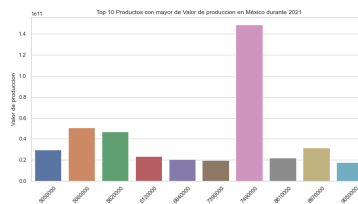


Figura 21: Top10 Productos Mexico 2021

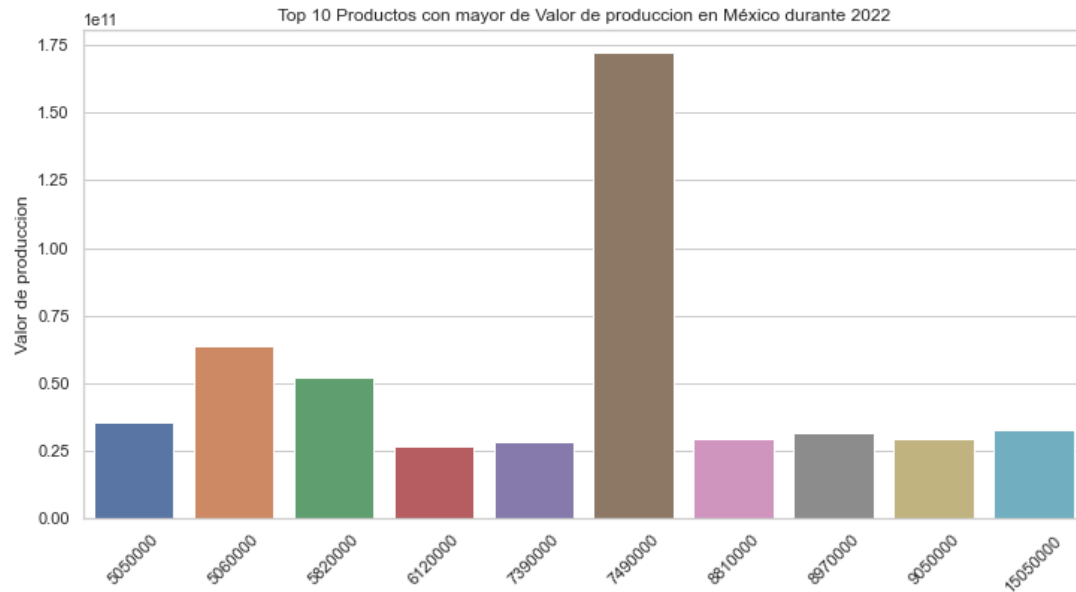


Figura 22: Top10 Productos Mexico 2022

Con base a la informacion que nos proporcionan estas graficas se pudo observar como los principales productos que tienen un mejor valor en el mercado son:

Tabla 1: Mejores productor en México

Codigo	Producto
7490000	Maíz grano
5820000	Chile verde
8810000	Sorgo grano
8970000	Tomate rojo
7390000	Limón

Además, al disponer de un sistema de tablas que facilitaba la unión y consulta eficiente de la información, se llevó a cabo un análisis similar, pero esta vez centrándose en el estado de Chihuahua.

A partir del DataFrame obtenido, se generaron gráficos básicos que representan el Top 10 de productos en el estado, los cuales generaron el mayor valor de producción en el mercado. A continuación, se presentan las visualizaciones resultantes.

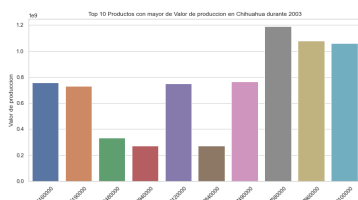


Figura 23: Top10 Productos Chihuahua 2003

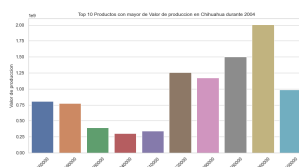


Figura 24: Top10 Productos Chihuahua 2004

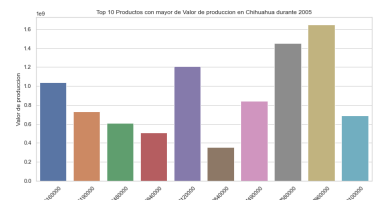


Figura 25: Top10 Productos Chihuahua 2005

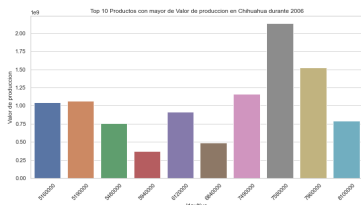


Figura 26: Top10 Productos Chihuahua 2006

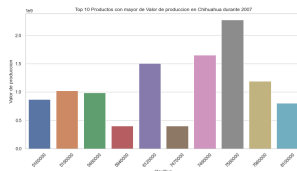


Figura 27: Top10 Productos Chihuahua 2007

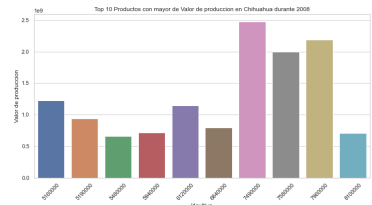


Figura 28: Top10 Productos Chihuahua 2008

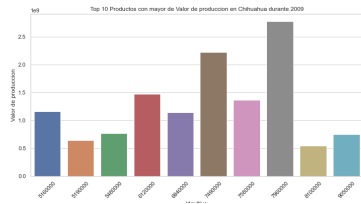


Figura 29: Top10 Productos Chihuahua 2009

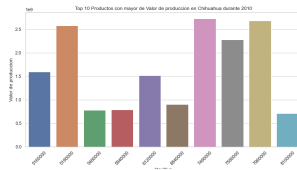


Figura 30: Top10 Productos Chihuahua 2010

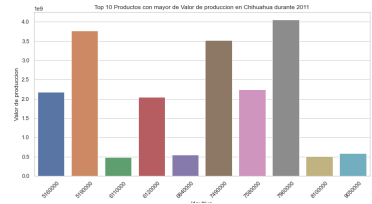


Figura 31: Top10 Productos Chihuahua 2011

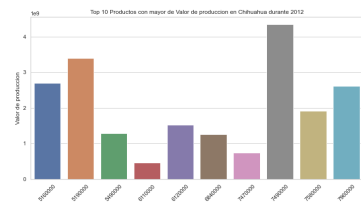


Figura 32: Top10 Productos Chihuahua 2012

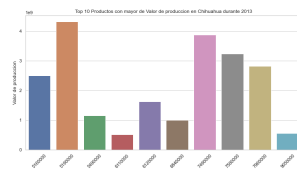


Figura 33: Top10 Productos Chihuahua 2013

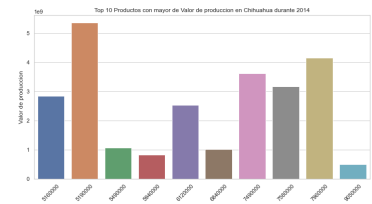


Figura 34: Top10 Productos Chihuahua 2014

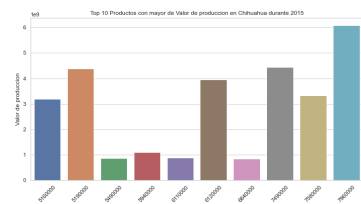


Figura 35: Top10 Productos Chihuahua 2015

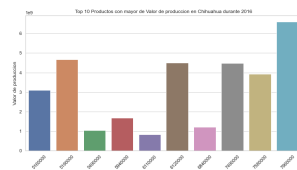


Figura 36: Top10 Productos Chihuahua 2016

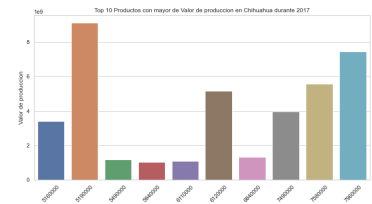


Figura 37: Top10 Productos Chihuahua 2017

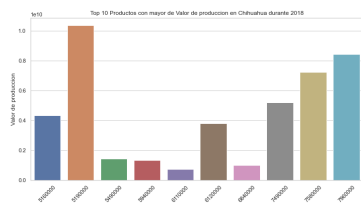


Figura 38: Top10 Productos Chihuahua 2018

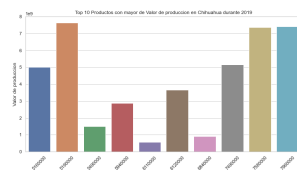


Figura 39: Top10 Productos Chihuahua 2019

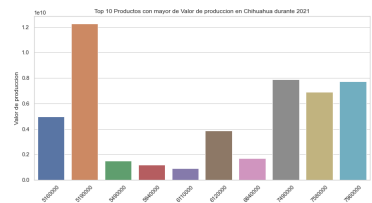


Figura 40: Top10 Productos Chihuahua 2021

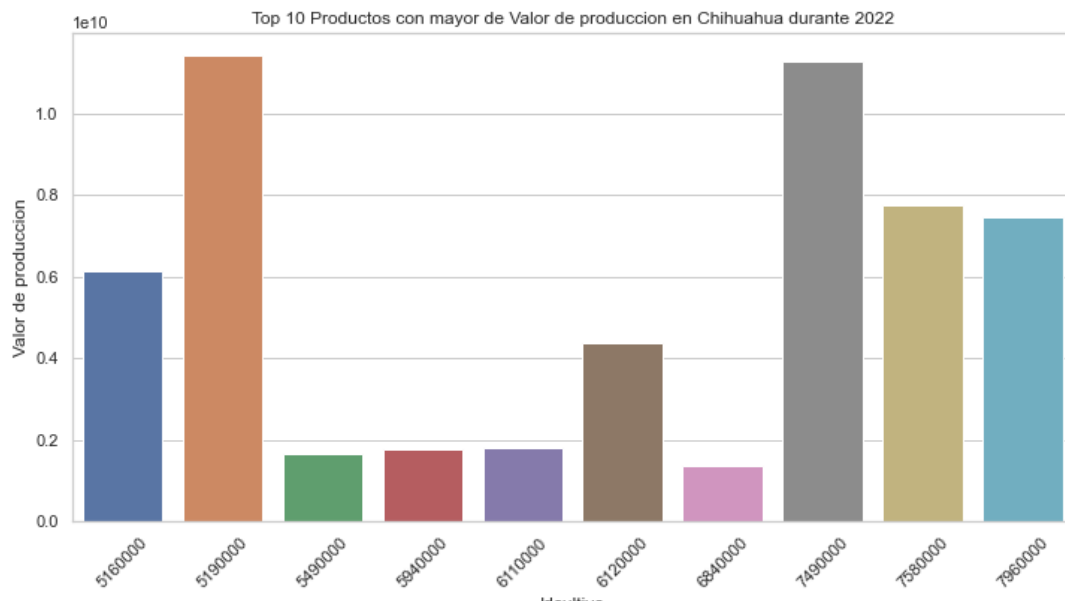


Figura 41: Top10 Productos Chihuahua 2022

Con base en la información proporcionada por estas gráficas, se observó que los principales productos que destacan por su mayor valor en el mercado son:

Tabla 2: Mejores productor en Chihuahua

Codigo	Producto
7490000	Maíz grano
6120000	Chile verde
5190000	Algodón hueso
5160000	Alfalfa achicalada
7060000	Guayaba

Y finalmente, se llevó a cabo el mismo proceso, pero esta vez se centró en Ciudad Juárez. Se observó que la Id de Ciudad Juárez variaba entre algunos archivos, por lo que se tuvieron en cuenta estas diferencias al realizar las consultas en la base de datos.

Con base en el dataframe obtenido, se generaron gráficas básicas mostrando el Top 10 de productos que generaron un mayor valor de producción en el mercado, resultando en las siguientes representaciones visuales.

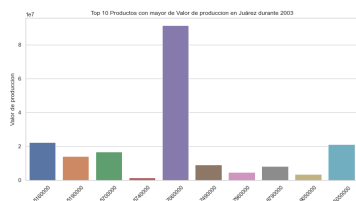
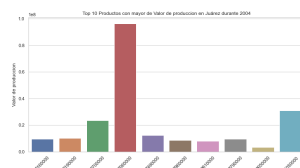


Figura 42: Top10 Productos Juarez 2003



2004

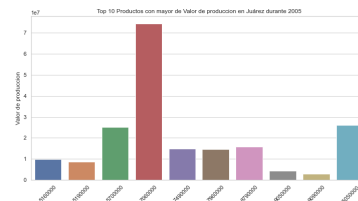


Figura 44: Top10 Productos Juarez 2005

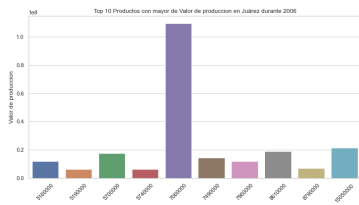


Figura 45: Top10 Productos Juarez 2006

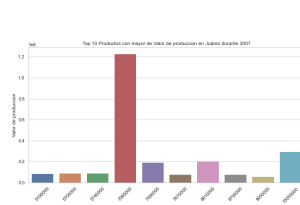


Figura 46: Top10 Productos Juarez 2007

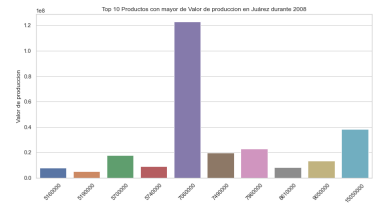


Figura 47: Top10 Productos Juarez 2008

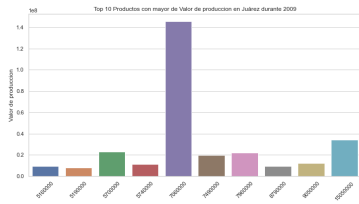


Figura 48: Top10 Productos Juarez 2009

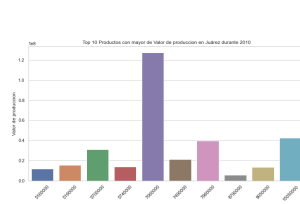


Figura 49: Top10 Productos Juarez 2010

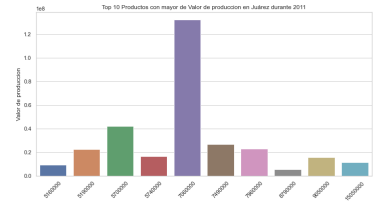


Figura 50: Top10 Productos Juarez 2011

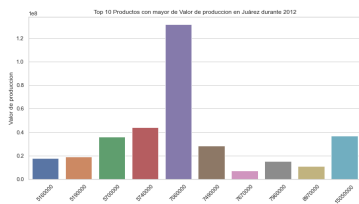


Figura 51: Top10 Productos Juarez 2012

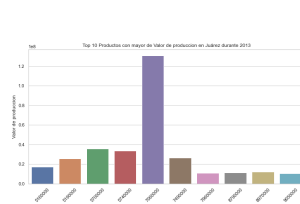


Figura 52: Top10 Productos Juarez 2013

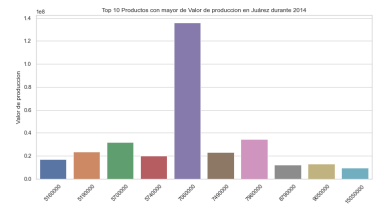


Figura 53: Top10 Productos Juarez 2014

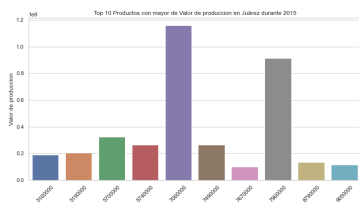


Figura 54: Top10 Productos Juarez 2015

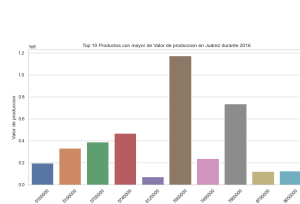


Figura 55: Top10 Productos Juarez 2016

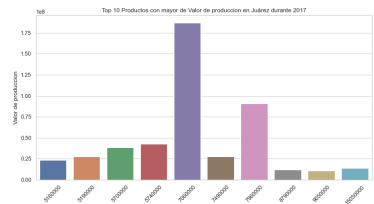


Figura 56: Top10 Productos Juarez 2017

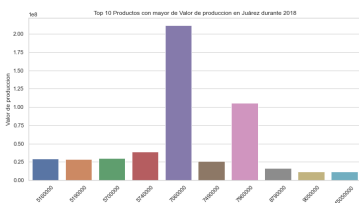


Figura 57: Top10 Productos Juarez 2018

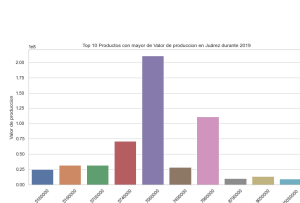


Figura 58: Top10 Productos Juarez 2019

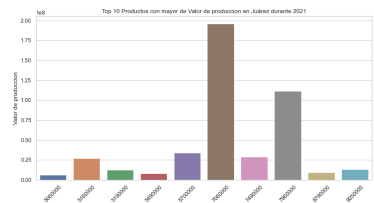


Figura 59: Top10 Productos Juarez 2021

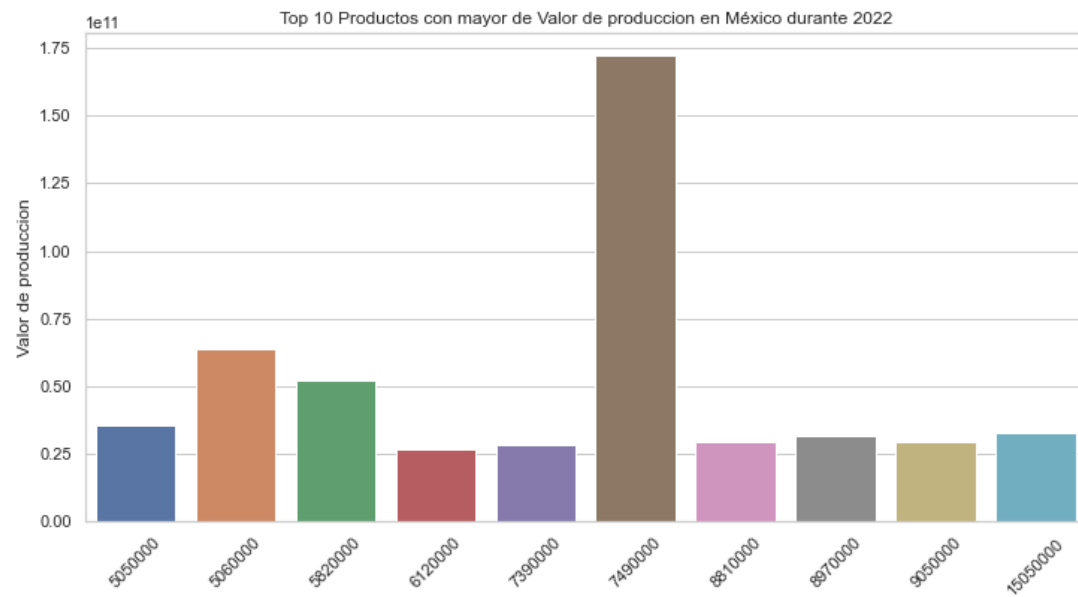


Figura 60: Top10 Productos Juarez 2022

Con base a la informacion que nos proporcionan estas graficas se pudo observar como los principales productos que tienen un mejor valor en el mercado son:

Tabla 3: Mejores productor en Juárez

Codigo	Producto
7490000	Maíz grano
6120000	Chile verde
8810000	Sorgo grano
8970000	Tomate rojo
7390000	Limón

En conclusión, este enfoque sistemático permitió identificar y visualizar los principales productos con un alto valor en el mercado, pero me he quedado corto en la presentación de los resultados.