

## ✓ Proyecto Bootcamp Analisis de Datos

Informacion delictiva del municipio de Bucaramanga - Detectar Variables y patrones para predecir y reducir la criminalidad

Presentado por Andres Camilo Lopez - Analista de Datos

CC. 1.053.807.414

### PROBLEMATICA

La creciente incidencia de delitos en diferentes barrios de Bucaramanga ha generado preocupación entre los residentes y las autoridades locales. La falta de un análisis inicial de las variables asociadas a la delincuencia (como los tipos de armas, horarios, ubicaciones y características de las víctimas) dificulta la implementación de políticas preventivas efectivas. Se necesita identificar patrones específicos que puedan ayudar a las autoridades a predecir y reducir la criminalidad en áreas de alta incidencia.

### OBJETIVO

- Analizar los patrones de conducta delictiva en Bucaramanga.
- Examinar las relaciones entre tipo de delito, horario, ubicación (barrio), edad y género de las personas afectadas, así como el tipo de arma utilizada.
- Proporcionar recomendaciones a las autoridades para la implementación de medidas de seguridad en los barrios más afectados y sugerir horarios de patrullaje.

### FUENTE DE DATOS

Información de los delitos ocurridos en el municipio según la modalidad y conducta delictiva, barrios y comunas de ocurrencia, armas y medios empleados, móvil del agresor y de la víctima, curso de vida y género de la víctima, con una desagregación temporal por mes, día, y hora de ocurrencia. de enero 2016 a octubre 2023

Se va a realizar el analisis de los datos de los años 2022 y 2023.

Se cuentan con 28116 registros donde cada fila es un una fila, un delito o crimen.

Los datos se tomaron de [https://www.datos.gov.co/Seguridad-y-Defensa/Informaci-n-delictiva-del-municipio-de-Bucaramanga/x46e-abhz/about\\_data](https://www.datos.gov.co/Seguridad-y-Defensa/Informaci-n-delictiva-del-municipio-de-Bucaramanga/x46e-abhz/about_data) datos abierto del gobierno nacional

## ✓ CRONOGRAMA

### ✓ 1. Importar librerias

```
1 #@title 1. Importar librerias
2
3 import pandas as pd
4 import plotly.express as px
```

### ✓ 2. Cargar Datos

```

1 #@title 2. Cargar Datos
2
3 datos = pd.read_excel('/content/drive/MyDrive/Información delictiva del municipio de Bucaramanga/informacion_delictiva_municip

```

### 3. Visualizar Metadatos

```

1 #@title 3. Visualizar Metadatos
2
3 #datos.head()
4 datos.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28116 entries, 0 to 28115
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   DESCRIPCION_CONDUCTA  28116 non-null  object
 1   ARMAS_MEDIOS          28116 non-null  object
 2   BARRIOS_HECHO         28116 non-null  object
 3   FECHA_HECHO           28116 non-null  datetime64[ns]
 4   HORA_HECHO            28116 non-null  object
 5   EDAD                  27483 non-null  object
 6   SEXO                  28116 non-null  object
 7   MOVIL_AGRESOR        28116 non-null  object
 8   CLASE_SITIO           28116 non-null  object
 9   AÑO_NUM               28116 non-null  int64
10  DIA_NUM               28116 non-null  int64
11  DIA_NOMBRE            28116 non-null  object
12  NOM_COM               28116 non-null  object
dtypes: datetime64[ns](1), int64(2), object(10)
memory usage: 2.8+ MB

```

```

1 ###@title Convertir Datos
2
3 #convertir el nombre del día en mayúscula
4 datos['DIA_NOMBRE'] = datos['DIA_NOMBRE'].str.upper()
5 #print(datos['DIA_NOMBRE'])

```

### 4. Limpiar Datos

```

1 #@title 4. Limpiar Datos
2 #consultar valores nulos
3 #print(datos[datos['EDAD'].isna()])
4 #se halla la media de edad y se reemplaza nulos y no disponible por 0
5 datos.loc[datos['EDAD'] == 'NO DISPONIBLE', 'EDAD'] = 0
6 datos.loc[datos['EDAD'] == 'NaN', 'EDAD'] = 0
7 datos['EDAD'] = datos['EDAD'].astype(float)
8 media = datos['EDAD'].mean()
9 #se reemplazan los valores nulos por la media - rellene los valores faltantes de edad y que realice el cambio en el dataframe
10 datos.fillna({'EDAD': media}, inplace=True)
11 print(datos[datos['EDAD'].isna()])
12 datos.describe()
13 #ahí un dato en la columna edad que es atípico max edad 125
14 #*****
15 #Valores duplicados
16 filas_duplicadas = datos[datos.duplicated()]
17 filas_duplicadas.to_excel('filas_duplicadas.xlsx', index=False)
18 print(filas_duplicadas)
19 datos.drop_duplicates(inplace=True)
20 filas_duplicadas = datos[datos.duplicated()]
21 datos.duplicated().sum()
22 print(filas_duplicadas)
23
24
25

```



25301	2023-10-10	21:50:00	61.0	FEMENINO	A PIE
25605	2023-10-24	00:00:00	41.0	FEMENINO	A PIE
26707	2023-07-05	13:40:00	29.0	MASCULINO	A PIE
26710	2023-07-05	13:40:00	34.0	MASCULINO	A PIE
28100	2023-10-20	14:30:00	31.0	FEMENINO	A PIE

	CLASE_SITIO	AÑO_NUM	DIA_NUM	DIA_NOMBRE	\
2321	VIAS PUBLICAS	2022	19	MIÉRCOLES	
3167	PARQUEADERO	2022	1	MARTES	
6133	DISCOTECAS	2022	20	VIERNES	
6872	VIAS PUBLICAS	2022	2	SÁBADO	
6973	VIAS PUBLICAS	2022	8	VIERNES	
7995	PARQUEADERO	2022	26	VIERNES	
10462	VIAS PUBLICAS	2022	30	MIÉRCOLES	
11881	VIAS PUBLICAS	2022	11	VIERNES	
13243	VIAS PUBLICAS	2022	31	MARTES	
14076	VIAS PUBLICAS	2022	4	DOMINGO	
15945	VIAS PUBLICAS	2023	8	SÁBADO	
16476	VIAS PUBLICAS	2023	6	LUNES	
18048	VIAS PUBLICAS	2023	14	JUEVES	
18050	VIAS PUBLICAS	2023	14	JUEVES	
18512	VIAS PUBLICAS	2023	18	MIÉRCOLES	
19086	VIAS PUBLICAS	2023	7	MARTES	
19861	VIAS PUBLICAS	2023	9	JUEVES	
20430	VIAS PUBLICAS	2023	1	SÁBADO	
20547	VIAS PUBLICAS	2023	7	VIERNES	
23469	VIAS PUBLICAS	2023	2	MIÉRCOLES	
25301	VIAS PUBLICAS	2023	10	MARTES	
25605	VIAS PUBLICAS	2023	24	MARTES	
26707	INSTALACIONES DE LA POLICIA	2023	5	MIÉRCOLES	
26710	INSTALACIONES DE LA POLICIA	2023	5	MIÉRCOLES	
28100	VIAS PUBLICAS	2023	20	VIERNES	

	NOM_COM
2321	NORORIENTAL
3167	PROVENZA
6133	OCCIDENTAL
6872	LA CIUDADELA
6973	CABECERA DEL LLANO
7995	ORIENTAL
10462	CABECERA DEL LLANO
11881	NORTE
13243	SAN FRANCISCO
14076	MORRORICO
15945	CENTRO
16476	SAN FRANCISCO
18048	CENTRO
18050	ORIENTAL
18512	ORIENTAL
19086	SAN FRANCISCO
19861	ORIENTAL
20430	SAN FRANCISCO
20547	SAN FRANCISCO

## 5. Filtrar datos

```

1 #@title 5. Filtrar datos
2 #filtrar po año, mes, dia , hora
3 datos['Año'] = datos['FECHA_HECHO'].dt.year
4 datos['Mes'] = datos['FECHA_HECHO'].dt.month_name()
5 datos['Dia'] = datos['FECHA_HECHO'].dt.dayofweek
6 #funcion para pasar dia a español dependiendo el numero del dia que arroje dayofweek
7 def dia_de_la_semana_en_espanol(dia_ingles):
8     dias_espanol = {0: 'Lunes', 1: 'Martes', 2: 'Miércoles', 3: 'Jueves', 4: 'Viernes', 5: 'Sábado', 6: 'Domingo'}
9     return dias_espanol[dia_ingles]
10 datos['Dia_español'] = datos['FECHA_HECHO'].dt.dayofweek.apply(lambda x: dia_de_la_semana_en_espanol(x))
11 datos['Hora'] = datos['FECHA_HECHO'].dt.hour
12 #conteo de registros del 2022 y 2023
13 conteo_2023 = datos[datos['Año'] == 2023].shape[0]
14 conteo_2022 = datos[datos['Año'] == 2022].shape[0]
15 datos
16

```



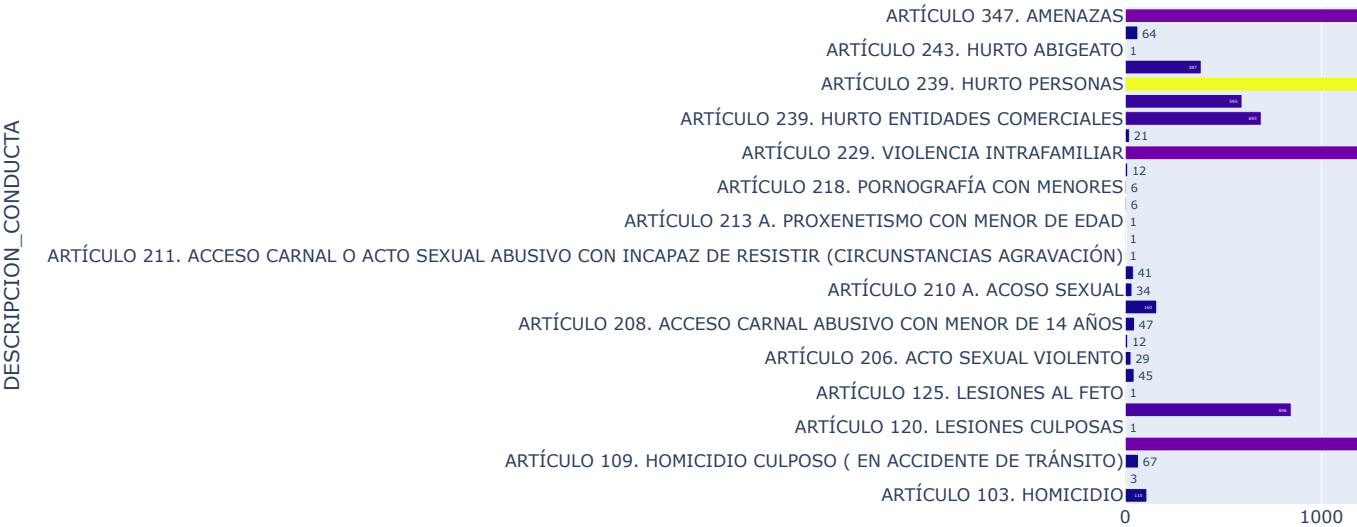
```

7         labels={'x': 'Frecuencia', 'y': 'Descripción de Conducta'},title='Frecuencia de Descripciones de Conducta',
8         text_auto=True,
9         color = conteo_por_descripcion_2022)
10
11 # Crear el gráfico de barras conteo por descripcion de conducta para 2023
12 fig_2023 = px.bar(conteo_por_descripcion_2023,
13                   x=conteo_por_descripcion_2023.values,
14                   y=conteo_por_descripcion_2023.index,
15                   labels={'x': 'Frecuencia', 'y': 'Descripción de Conducta'},title='Frecuencia de Descripciones de Conducta',
16                   text_auto=True,
17                   template='plotly_dark',
18                   color = conteo_por_descripcion_2023)
19
20 # grafico de barras verticales armas o medios utilizados
21 fig_armas = px.bar(conteo_por_armas,
22                   x=conteo_por_armas.index,
23                   y=conteo_por_armas.values,
24                   labels={'x': 'Arma', 'y': 'Frecuencia'},
25                   title='Armas utilizadas para los delitos',
26                   text_auto=True,
27                   color = conteo_por_armas)
28
29 # Mostrar los gráficos
30 fig_2022.show()
31 fig_2023.show()
32 fig_armas.show()
33

```



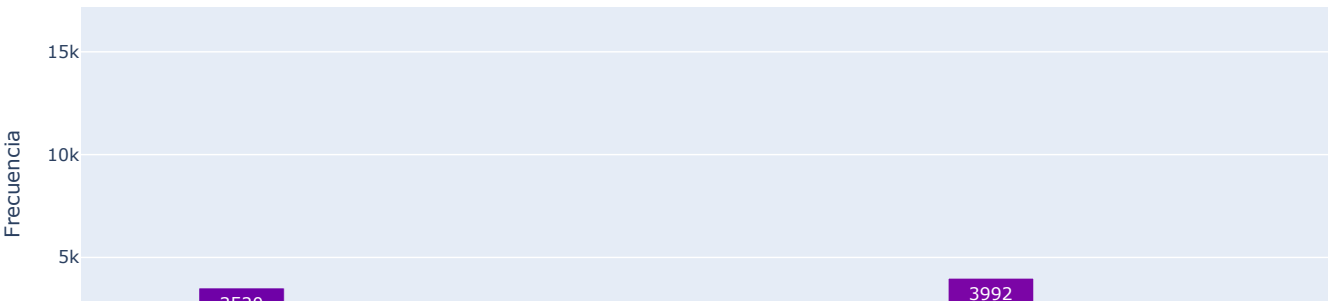
Frecuencia de Descripciones de Conducta

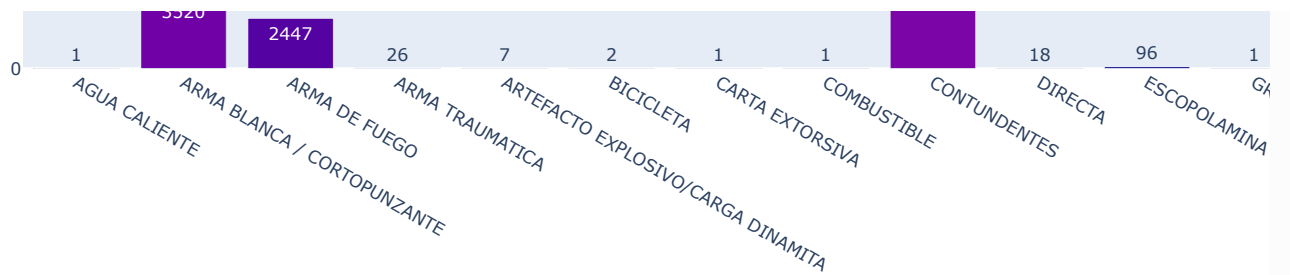


Frecuencia de Descripciones de Conducta



Armas utilizadas para los delitos





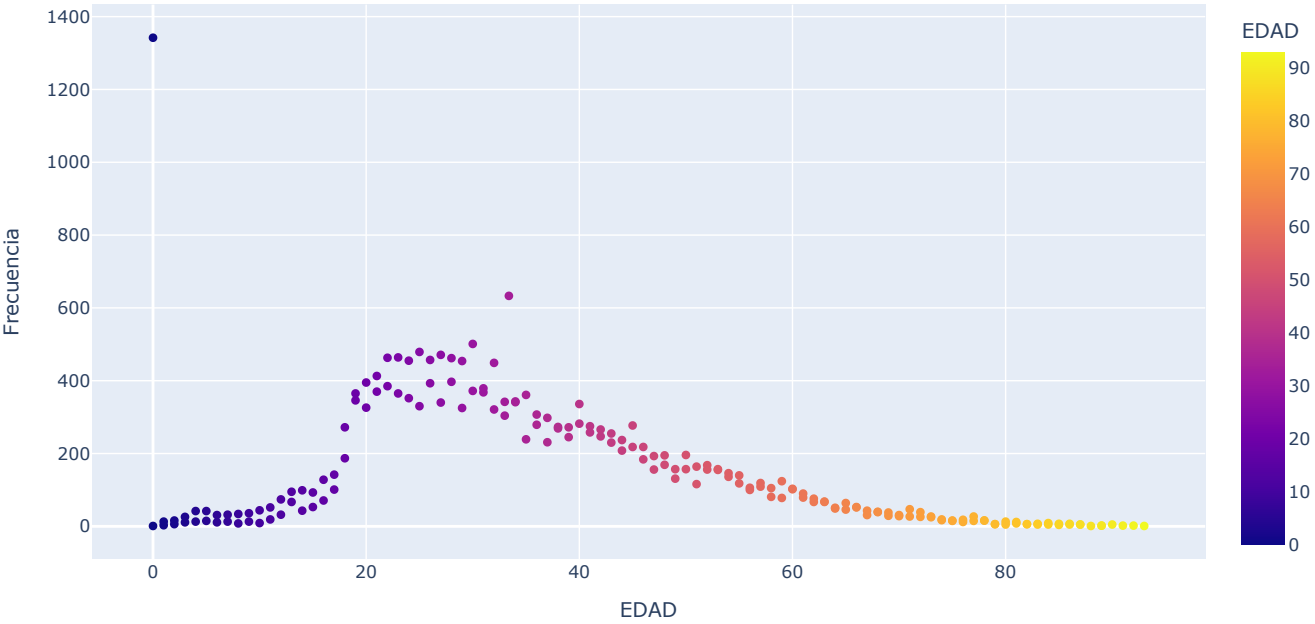
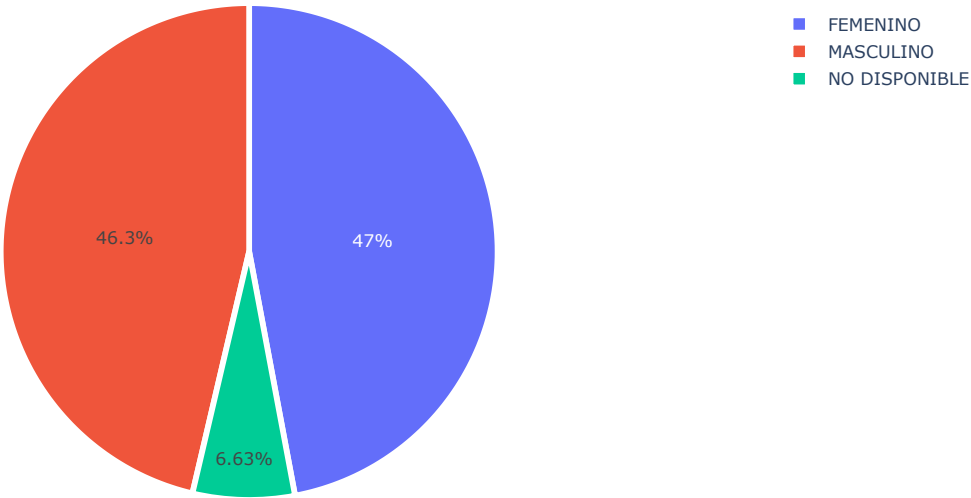
```

1 # grafico tipo pie genero de la victima de los delitos
2 fig_genero = px.pie(genero,
3     values=genero.values,
4     names=genero.index,
5     labels={'x': 'Genero', 'y': 'Frecuencia'},
6     title='Genero victima de los delitos',)
7 #dividir torta
8 fig_genero.update_traces(marker = dict(line = dict(color = 'white', width = 4)))
9
10 #Distribución de la edad de las víctimas
11 fig_edad = px.scatter(df_combined, x='EDAD', y='Frecuencia', color='EDAD')
12
13 #tendencia por mes 2022 - 2023
14 fig_tendencia_2022 = px.line(tendencia_2022,
15     x=tendencia_2022.index,
16     y=tendencia_2022.values,
17     labels={'x': 'Mes', 'y': 'Frecuencia'},
18     title= 'Delitos por Mes')
19
20 fig_tendencia_2022.update_traces(line = dict(dash = "dot", width = 4, color = "red"),
21     marker = dict(color = "darkblue", size = 20, opacity = 0.8))
22
23 fig_tendencia_2023 = px.line(tendencia_2023,
24     x=tendencia_2023.index,
25     y=tendencia_2023.values,
26     labels={'x': 'Mes', 'y': 'Frecuencia'},
27     title= 'Delitos por Mes')
28
29 fig_tendencia_2023.update_traces(line = dict(dash = "dot", width = 4, color = "blue"),
30     marker = dict(color = "darkblue", size = 20, opacity = 0.8))
31
32
33 # Mostrar los gráficos
34
35 fig_genero.show()
36 fig_edad.show()
37 fig_tendencia_2022.show()
38 fig_tendencia_2023.show()

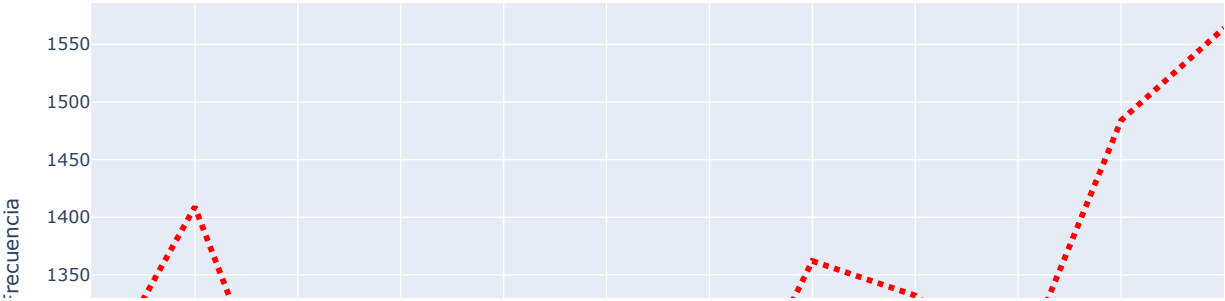
```



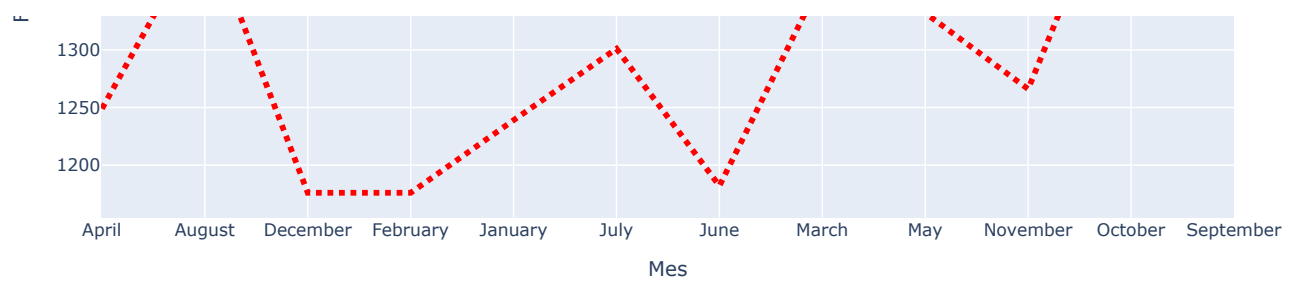
Genero victima de los delitos



Delitos por Mes







### Delitos por Mes

