

Factores causantes en los accidentes y la criminalidad

-
- DANIEL ANDRES BECERRA SIERRA.
 - JUAN JOSE RODRIGUEZ
 - Jhonatan Rodrigo Robayo Barrera



Entendimiento del negocio

Nueva York con una población diversa posee una población de aproximadamente 8.468 millones de habitantes.

Crecimiento del empleo: Tuvo un aumento del 5% en 2022

Tasa de desempleo: En noviembre de 2023 la tasa de desempleo fue del 4,3%

Crecimiento del PIB : En 2023 el crecimiento del PIB fue del 3,5%

Crecimiento de ingresos personales: crecimiento del 0.8%

Disminución de la población: en 2022 hubo una disminución de la población del 0.4%



Entendimiento del negocio

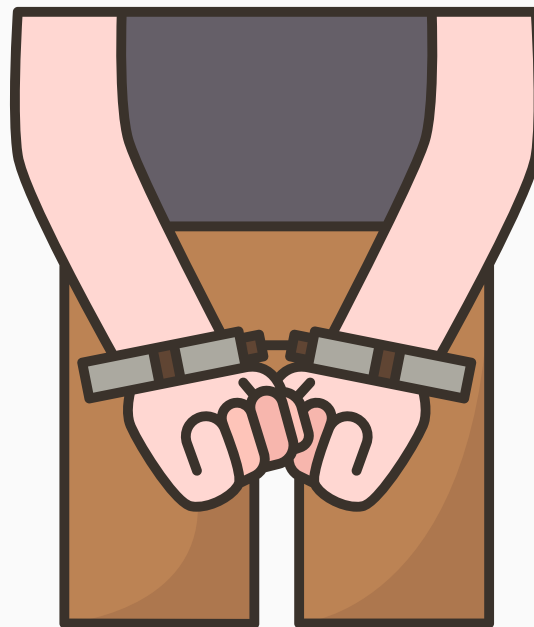
Objetivo

Identificar diferentes factores detonantes o relacionales específicos de accidentes y los principales casos de delitos además de las relaciones potenciales entre diferentes datos demográficos y de pobreza con estos eventos, para así poder desarrollar un plan de acción detallado y basado en análisis de datos que permita al estado de Nueva York abordar estos problemas de manera efectiva y permitiendo la reducción de los mismos.



Selección de los datos

NYPD Arrest Data (Year to Date)



Motor Vehicle Collisions - Vehicles



Coleccion y Revision de Datos

Arrest

PERP_RACE: string (nullable = true) Raza del perpetrador

X_COORD_CD: integer (nullable = true) Coordenadas de la ubicación del arresto en el Sistema de Coordenadas Planas del Estado de Nueva York

Y_COORD_CD: integer (nullable = true) Coordenadas de la ubicación del arresto en el Sistema de Coordenadas Planas del Estado de Nueva York

Latitude: double (nullable = true) Coordenadas de latitud para el sistema de coordenadas global

Longitude: double (nullable = true) Coordenadas de longitud para el sistema de coordenadas global

New Georeferenced Column: string (nullable = true) Columna que combina la latitud y longitud en un solo punto georreferenciado

ARREST_KEY: integer (nullable = true). Identificador único y persistente generado aleatoriamente para cada arresto.

ARREST_DATE: date (nullable = true) Fecha exacta del arresto.

PD_CD: integer (nullable = true) Código de clasificación interna de tres dígitos, más detallado que el código KY. Ayuda a identificar el tipo específico de delito por el cual se efectuó el arresto.

PD_DESC: string (nullable = true) Descripción de la clasificación interna correspondiente al código PD, proporcionando detalles más granulares sobre el delito.

KY_CD: integer (nullable = true) Código de clasificación interna de tres dígitos que indica una categoría más general de delito que el código PD.

OFNS_DESC: string (nullable = true) Descripción de la clasificación interna correspondiente al código KY, indicando la categoría más general del delito

LAW_CODE: string (nullable = true) Códigos de la ley bajo los cuales se efectuaron los cargos, incluyendo la Ley Penal de NYS, VTL (Ley de Tránsito Vehicular) y otras leyes locales.

LAW_CAT_CD: string (nullable = true) Nivel del delito (felonía, delito menor, infracción)

ARREST_BORO: string (nullable = true) Abreviatura del borough de Nueva York donde se realizó el arresto (Bronx, Staten Island, Brooklyn, Manhattan, Queens).

ARREST_PRECINCT: integer (nullable = true) Precinto donde ocurrió el arresto

JURISDICTION_CODE: integer (nullable = true) Código de la jurisdicción responsable del arresto, donde los códigos 0 (Patrulla), 1 (Tránsito) y 2 (Vivienda) representan jurisdicciones del NYPD, y códigos 3 en adelante representan otras jurisdicciones.

AGE_GROUP: string (nullable = true) Grupo de edad del perpetrador

PERP_SEX: string (nullable = true) Grupo de edad del perpetrador

Coleccion y Revision de Datos

Collisions

VEHICLE_DAMAGE_1: string (nullable = true). Ubicaciones adicionales del daño en el vehículo.

VEHICLE_DAMAGE_2: string (nullable = true). Ubicaciones adicionales del daño en el vehículo.

VEHICLE_DAMAGE_3: string (nullable = true). Ubicaciones adicionales del daño en el vehículo.

PUBLIC_PROPERTY_DAMAGE: string (nullable = true). Daño a propiedad pública (Sí o No).

PUBLIC_PROPERTY_DAMAGE_TYPE: string (nullable = true). Tipo de propiedad pública dañada (por ejemplo, señal, cerca, poste de luz, etc.).

CONTRIBUTING_FACTOR_1: string (nullable = true). Factores que contribuyen a la colisión para el vehículo designado.

CONTRIBUTING_FACTOR_2: string (nullable = true). Factores adicionales que contribuyen a la colisión para el vehículo designado.

UNIQUE_ID: integer (nullable = true). Código único de registro generado por el sistema, que sirve como clave primaria.

COLLISION_ID: integer (nullable = true). Código de identificación del choque. Clave foránea que coincide con unique_id de la tabla de choques.

CRASH_DATE: date (nullable = true). Fecha de ocurrencia de la colisión.

CRASH_TIME: timestamp (nullable = true). Hora de ocurrencia de la colisión.

VEHICLE_ID: string (nullable = true). Código de identificación del vehículo asignado por el sistema.

STATE_REGISTRATION: string (nullable = true). Estado donde el vehículo está registrado.

VEHICLE_TYPE: string (nullable = true). Tipo de vehículo basado en la categoría de vehículo seleccionada (ATV, bicicleta, carro/suv, ebike, scooter, camión/bus, motocicleta, otro).

VEHICLE_MAKE: string (nullable = true). Marca del vehículo.

VEHICLE_MODEL: string (nullable = true). Modelo del vehículo.

VEHICLE_YEAR: integer (nullable = true). Año de fabricación del vehículo.

TRAVEL_DIRECTION: string (nullable = true). Dirección hacia la cual viajaba el vehículo.

VEHICLE_OCCUPANTS: integer (nullable = true). Número de ocupantes en el vehículo.

DRIVER_SEX: string (nullable = true). Sexo del conductor.

DRIVER_LICENSE_STATUS: string (nullable = true). Estado de la licencia del conductor (licencia, permiso, no licenciado).

DRIVER_LICENSE_JURISDICTION: string (nullable = true). Estado donde fue emitida la licencia del conductor.

PRE_CRASH: string (nullable = true). Acción previa al choque (ir recto, girar a la derecha, adelantar, retroceder, etc.).

POINT_OF_IMPACT: string (nullable = true). Ubicación en el vehículo del punto inicial de impacto (por ejemplo, lado del conductor, lado trasero del pasajero, etc.).

VEHICLE_DAMAGE: string (nullable = true). Ubicación en el vehículo donde ocurrió la mayor parte del daño.

	summary ▲	ARREST_KEY ▲	PD_CD ▲	PD_DESC ▲	KY_CD ▲	OFNS_DESC ▲	LAW_CODE ▲
1	count	226872	226870	226872	226855	226872	226872
2	mean	2.706479248400111E8	424.7544011989245	null	249.3451323532653	null	null
3	stddev	5304010.298148577	274.4753806048603	null	147.68673264760508	null	null
4	min	261180920	1	(null)	101	(null)	(null)
5	max	279779734	997	WEAPONS,MFR,TRANSPORT,ETC.	995	VEHICLE AND TRAFFIC LAWS	VTL21300A5

magen 1.1 recorte muestra de datos

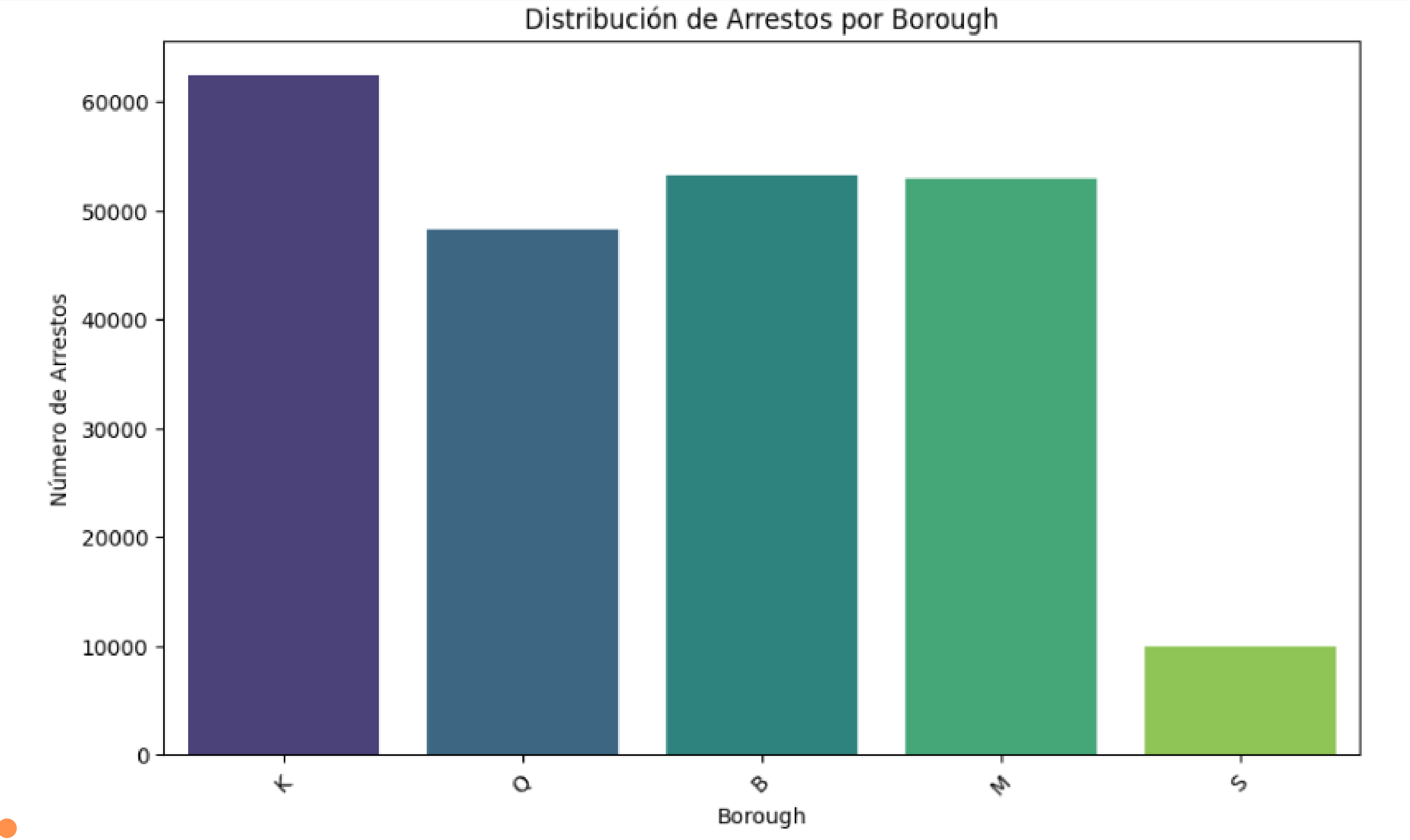
LAW_CAT_CD ▲	ARREST_BORO ▲	ARREST_PRECINCT ▲	JURISDICTION_CODE ▲	AGE_GROUP ▲	PERP_SEX ▲	PERP_RACE ▲	X_COORD_CD
225273	226872	226872	226872	226872	226872	226872	226872
9.0	null	63.43052910892486	0.9285367960788462	null	null	null	1005786.72871927
0.0	null	34.635045257003966	7.538568508006557	null	null	null	21509.4376481518
(null)	B	1	0	18-24	F	AMERICAN INDIAN/ALASKAN NATIVE	0
V	S	123	97	<18	U	WHITE HISPANIC	1067220

X_COORD_CD ▲	Y_COORD_CD ▲	Latitude ▲	Longitude ▲	New Georeferenced Column ▲
226872	226872	226872	226872	226872
1005786.7287192778	208289.0843206742	40.7381536574416	-73.92191484770285	null
21509.437648151856	29744.7188726473	0.1182365542455639	0.17333780170454163	null
0	0	0.0	-74.253256	POINT (-73.70059684703173 40.7390218775969)
1067220	271819	40.912714	0.0	POINT (0 0)

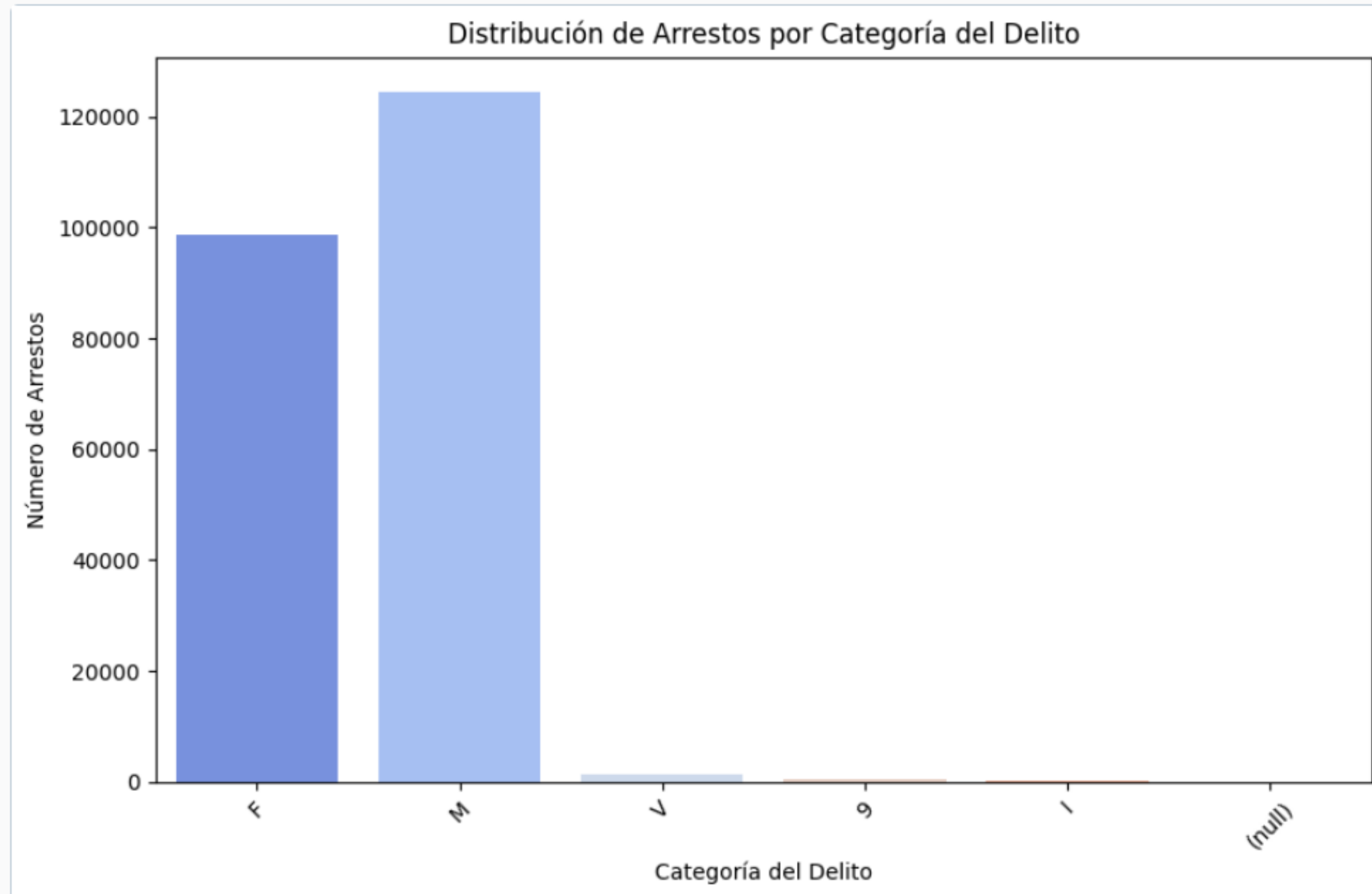
Concentración de arrestos es Nueva York



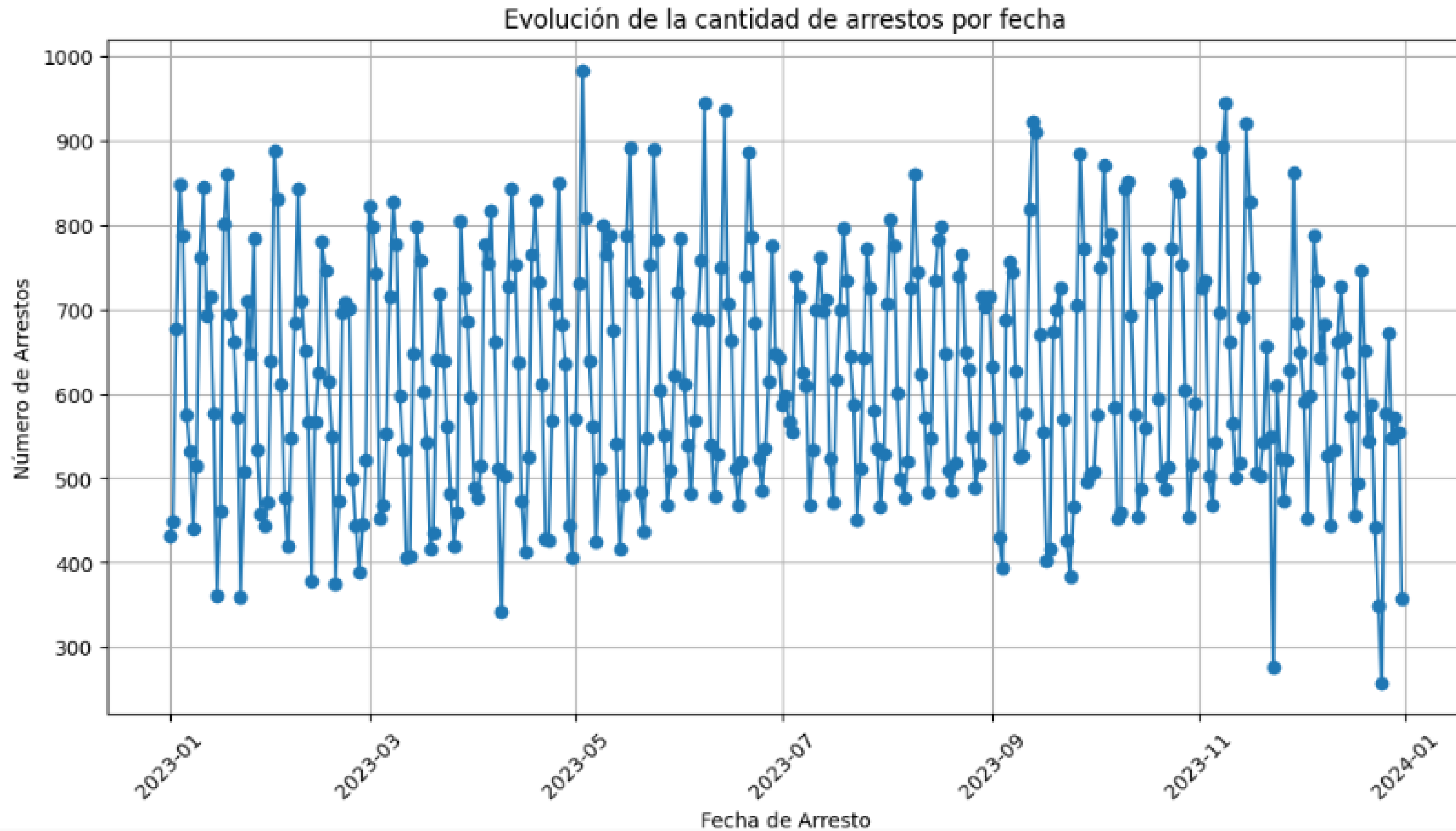
Arrestos por Borough



Arrestos por categoría de delito

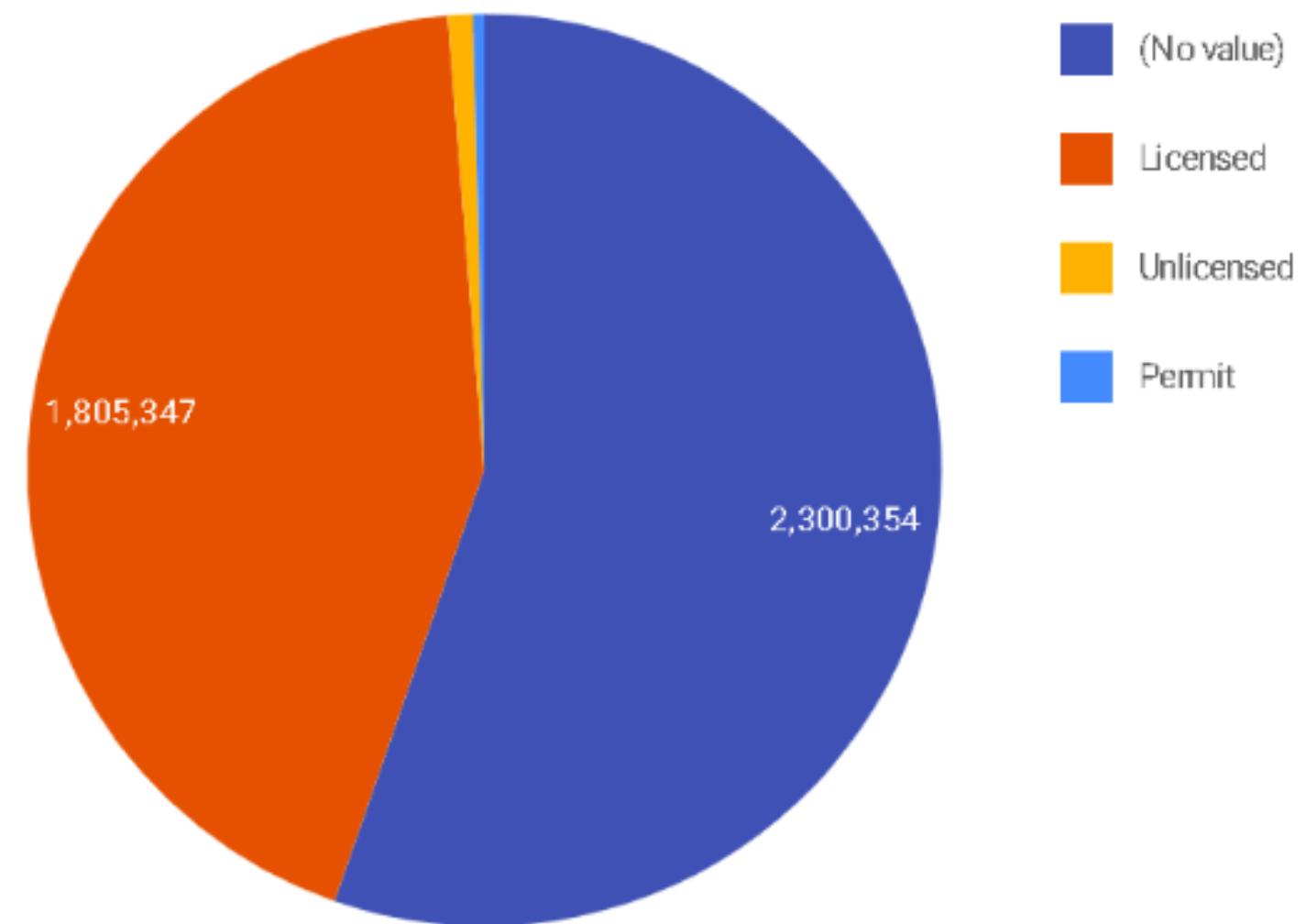


Arrestos en el tiempo





Accidentes en el Tiempo



Reporte de la calidad de los datos

NYPD Arrest Data (Year to Date)

Valores faltantes por columna:	
ARREST_KEY	0
ARREST_DATE	0
PD_CD	2
PD_DESC	0
KY_CD	17
OFNS_DESC	0
LAW_CODE	0
LAW_CAT_CD	1599
ARREST_BORO	0
ARREST_PRECINCT	0
JURISDICTION_CODE	0
AGE_GROUP	0
PERP_SEX	0
PERP_RACE	0
X_COORD_CD	0
Y_COORD_CD	0
Latitude	0
Longitude	0
New Georeferenced Column	0
dtype: int64	

OFNS_DESC
OTHER TRAFFIC INFRACTION
FELONY SEX CRIMES
OTHER OFFENSES RELATED TO THEF
VEHICLE AND TRAFFIC LAWS
KIDNAPPING & RELATED OFFENSES
OFF. AGNST PUB ORD SENSBLTY &
FELONY ASSAULT
ALCOHOLIC BEVERAGE CONTROL LAWS



Reporte de la calidad de los datos

Motor Vehicle Collisions – Vehicles

UNIQUE_ID	0
COLLISION_ID	0
CRASH_DATE	0
CRASH_TIME	0
VEHICLE_ID	0
STATE_REGISTRATION	299164
VEHICLE_TYPE	232911
VEHICLE_MAKE	1874569
VEHICLE_MODEL	4098613
VEHICLE_YEAR	1893963
TRAVEL_DIRECTION	1665649
VEHICLE_OCCUPANTS	1778286
DRIVER_SEX	2209212
DRIVER_LICENSE_STATUS	2296864
DRIVER_LICENSE_JURISDICTION	2291944
PRE_CRASH	918633
POINT_OF_IMPACT	1698458
VEHICLE_DAMAGE	1722396
VEHICLE_DAMAGE_1	2587669
VEHICLE_DAMAGE_2	2974293
VEHICLE_DAMAGE_3	3249573

PUBLIC_PROPERTY_DAMAGE	1528858
PUBLIC_PROPERTY_DAMAGE_TYPE	4124336
CONTRIBUTING_FACTOR_1	146119
CONTRIBUTING_FACTOR_2	1685347
dtype: int64	



Planteamiento de preguntas sobre los datos

1. ¿Cuáles serán las zonas de Nueva York con las tasas más altas de arrestos por cada 10,000 habitantes?
2. ¿Qué relación se presenta en cuanto a los arrestos por zona geográfica por conducción o accidente?
3. ¿Qué áreas con características similares en cuanto a tasas de arresto requieren de una mayor intervención?
4. ¿Cuál es el impacto de variables como la edad, ubicación y raza en la probabilidad de ser arrestado por un tipo específico de crimen?
5. ¿En qué ubicaciones hay una mayor probabilidad de que haya un responsable del arresto específico?.
6. ¿Existe una diferencia notable en el número de arrestos entre los diferentes días de la semana en Nueva York?
7. ¿Cuales son los días entre semana que más choques se presentaron, y los rangos de horas?
8. ¿Qué tipo de vehículos presentan más accidentes, cuantos ocupantes llevan en promedio?
9. ¿Cuáles serán las zonas con mayor cantidad de colisiones en para el próximo año basado en tendencias históricas?
10. ¿Qué factores contribuyen mas a un accidente basado en el clima frecuente por ese tiempo?



Filtros, limpieza y transformación inicial

NYPD Arrest Data (Year to Date)

Limpieza de columnas: Se eliminan múltiples columnas consideradas no relevantes para el análisis futuro.

Códigos y descripciones específicas del arresto (PD_CD, KY_CD, PD_DESC).

Motor Vehicle Collisions - Vehicles

Imputación de datos

VEHICLE_OCCUPANTS: por la mediana de la columna.

CONTRIBUTING_FACTOR_1: por la moda de la columna



Referencias

Bibliografía

[1] Brookings. (s/f). The Unequal Burden of Crime and Incarceration on America's Poor. Brookings Institution. Recuperado el 2 de febrero de 2024, de <https://www.brookings.edu/articles/the-unequal-burden-of-crime-and-incarceration-on-americas-poor/>

[2] Oficina del Controlador de Nueva York (OSC), O. del C. del E. (s/f). Economic and Demographic Trends. Oficina del Contralor Del Estado de Nueva York (OSC). Recuperado el 2 de febrero de 2024, de <https://www.osc.ny.gov/reports/finance/2023-fcr/economic-and-demographic-trends>

[3] Instituto de Investigación en Pobreza (IRP) de la Universidad de Wisconsin-Madison. (s/f). Connections Among Poverty, Incarceration, and Inequality. Instituto de Investigación En Pobreza (IRP) de La Universidad de Wisconsin-Madison. Recuperado el 2 de febrero de 2024, de <https://www.irp.wisc.edu/resource/connections-among-poverty-incarceration-and-inequality/>