

Procesamiento de Datos A Gran Escala:

Colisiones & Arrestos II

Por Daniela Torres, Daniel Sandoval e Isaac Janica

Índice

01

**DATA VIZ y Filtros
finales**

02

ML

03

Conclusiones

The background is a collage of torn, aged paper with various fragments of text like "On k...", "n...", "abs...", and "6...". A red line with four circular pins (two yellow, two red) is drawn across the collage, forming a frame around the central text area.

01

Data Viz y Filtros Finales

Dashboards interactivos para una mejor comprensión

Fue hecho en **Tableau**

Se va a hacer ingreso a tableau pero en el informe hay screenshots que muestran ejemplos de las gráficas y cómo cambian conforme a los parámetros



Fue hecho en **Tableau**

[https://public.tableau.com/app/profile/isaac.
janica5080/viz/DashBoardColisiones/Dashbo
ard1?publish=yes](https://public.tableau.com/app/profile/isaac.janica5080/viz/DashBoardColisiones/Dashbo
ard1?publish=yes)

[DashBoardColisiones | Tableau Public](#)

[DashBoardDelitos | Tableau Public](#)



+ a b l e a u



02

Machine Learning

Random Forest y K-Means

Random Forest

Predicción Hurto

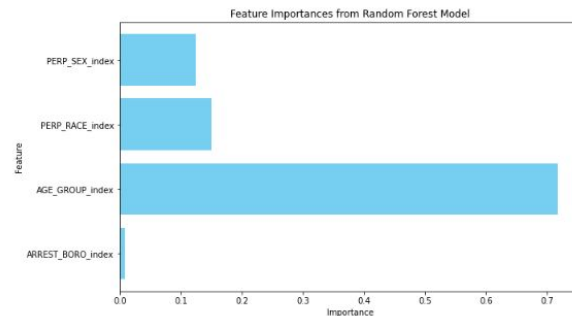
```
data: pyspark.sql.dataframe.DataFrame
```

```
ARREST_BORO_index: double  
AGE_GROUP_index: double  
PERP_RACE_index: double  
PERP_SEX_index: double  
Robbery_Binary: integer
```

```
# Split the data into training and testing sets (80% training, 20% testing)  
train_data, test_data = data.randomSplit([0.8, 0.2], seed=42)
```

```
importance_df: pyspark.sql.dataframe.DataFrame = [Feature: string, Importance: double]
```

Feature	Importance
ARREST_BORO_index	0.008249756037995253
AGE_GROUP_index	0.7182961778803223
PERP_RACE_index	0.14956078645580956
PERP_SEX_index	0.12389327962587295



Accuracy: 0.6774415906005872

Random Forest

Predicción hurto mayor

```
+-----+-----+
|           Feature|           Importance|
+-----+-----+
| ARREST_BORO_index| 0.07197880342248002|
| AGE_GROUP_index| 0.0698596402142934|
| PERP_RACE_index| 0.016618526060182204|
| PERP_SEX_index| 0.04019815108182278|
| ARREST_PRECINCT_i...| 0.8013448792212217|
+-----+-----+
Accuracy: 0.63680643047272
```

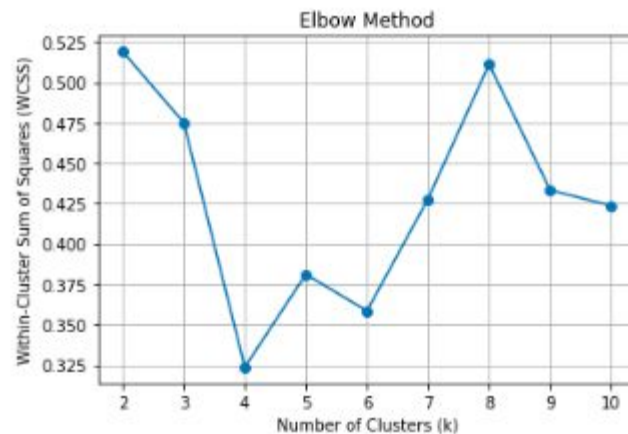
Predicción si raza
perpetrador es negra

```
+-----+-----+
|           Feature|           Importance|
+-----+-----+
| ARREST_BORO_index| 0.18409651741806976|
| OFNS_DESC_index| 0.010737294455013716|
| ARREST_PRECINCT_i...| 0.8051661881269165|
+-----+-----+
Accuracy: 0.641106264254556
Precision: 0.6406926300932817
Recall: 0.641106264254556
F1 Score: 0.6405882332556421
AUC: 0.7041555013829053
```


K-Means

```
df: pyspark.sql.dataframe.DataFrame  
  ARREST_BORO_index: float  
  AGE_GROUP_index: float  
  PERP_RACE_index: float  
  PERP_SEX_index: float
```

predictions: pyspark.sql.dataframe.DataFrame



Conclusiones Arrestos

- Ciudades como Nueva York, con economías e ingresos elevados, no están exentas de problemas de seguridad y accidentes viales. Aunque las métricas de seguridad muestran mejoras, no todos los problemas han sido completamente resueltos.
- Los delitos más comunes en NY son los asaltos de tipo 3 y ofensas relacionadas, el hurto menor y los delitos graves de asalto. Los días con mayor incidencia de crímenes son los miércoles, jueves y martes.
- Aunque la raza y el género pueden influir en la incidencia de crímenes, no se puede concluir precipitadamente que las personas de raza negra y los hombres son los principales responsables. Para ello sería necesario disponer de datos de enjuiciamiento para verificar qué proporción de estos grupos se declara culpable.
- La mayoría de los crímenes son cometidos por personas entre los 25 y 44 años, personas en un buenas condiciones físicas.

Conclusiones Colisiones

- La mayoría de los accidentes en NY son causados por residentes locales o de áreas cercanas, sin relación con personas de otros estados.
- Los vehículos más involucrados en accidentes son los sedanes y las camionetas SUV. En colisiones sin daños, bicicletas y box trucks.
- Los datos en colisiones muestran una reducción abrupta en la cantidad de accidentes después de pandemia, lo cual sería una buena pregunta de investigación.
- Los días con más incidentes son los jueves y viernes, sugiriendo una posible relación con la vida nocturna de la ciudad, aunque se requiere de una investigación más profunda.

Gracias!

MÁS CUIDADO RUFIÁN



NO VAYAS A MATAR
A ALGUIEN DE LA RISA



le dije ola i me
dijo cusbai