



Cryme Analysis for Bikers



Equipo Plata



Crime analysis for bikers

Equipo Plata ●



Eric Flores



**Alejandra
Pedroza**



Iván Galaviz



**Rogelio
Servín**



Hugo Alonzo

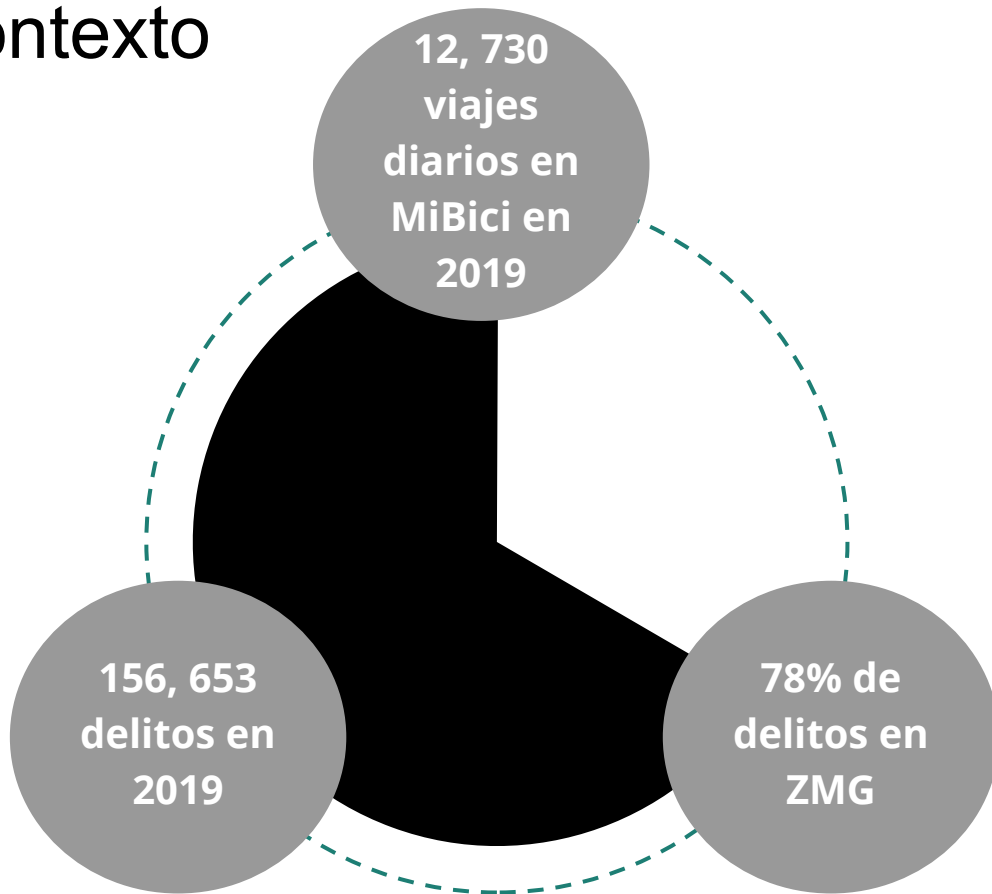


Mentor:

**Manuel David
Morales**



Contexto



Antecedentes

Encuesta “Usuarios del Sistema de Bicicletas Públicas en el Área Metropolitana de Guadalajara y su experiencia en Movilidad” por Kaliopéo.



Problema

Actualmente no existe una forma para determinar cuáles zonas de las que cubre el sistema MiBici son más seguras para los ciudadanos que utilizan dicho medio de transporte.



¿Por qué la Inteligencia Artificial es útil para este problema?

- Para estimar el impacto de los delitos
- Para pronosticar la ocurrencia de delitos



Decisiones de autoridades



Objetivos

- Estimar la posible relación entre las zonas con incidencia delictiva y las zonas que frecuentan los usuarios de MiBici.
- Pronosticar si los usuarios de MiBici pueden ser posibles víctimas de algún tipo de delito, de acuerdo a las zonas.



Hipótesis

- La incidencia delictiva afecta las zonas que los ciclistas frecuentan.
- Los usuarios de MiBici son posibles víctimas de delitos específicos, según las estaciones que utilizan.

Datos

Incidencia Delictiva en Jalisco, dataset del IIEG

- 2018 y 2019
- 98 mil 791 registros

	Año	Número_mes	Mes	Clave_Mun	Municipio	Localidad	Colonia	Delito	Cantidad
0	2018	1	Enero	1	ACATIC	CABECERA MUNICIPAL	LA VILLA	ROBO CASA HABITACION	1
1	2018	1	Enero	1	ACATIC	CABECERA MUNICIPAL	AGUA BLANCA	VIOLENCIA FAMILIAR	1
2	2018	1	Enero	1	ACATIC	CABECERA MUNICIPAL	CENTRO CAB. MUN.	LESIONES DOLOSAS	1
3	2018	1	Enero	1	ACATIC	CABECERA MUNICIPAL	CENTRO CAB. MUN.	VIOLENCIA FAMILIAR	1
4	2018	1	Enero	1	ACATIC	TIERRAS COLORADAS	N.D.	ROBO DE MOTOCICLETA	1

Viajes de MiBici, dataset del Sistema MiBici

	Viaje_Id	Usuario_Id	Genero	Anio_de_nacimiento	Inicio_del_viaje	Fin_del_viaje	Origen_Id	Destino_Id
8056494	8185641	36163	M	1984	2018-10-31 23:58:33	2018-11-01 00:12:12	60	180
8056495	8185642	10127	F	1994	2018-10-31 23:58:36	2018-11-01 00:25:26	53	53
8056496	8185643	28924	M	1999	2018-10-31 23:59:27	2018-11-01 00:10:16	79	177
8056497	8185644	14448	M	1993	2018-10-31 23:59:27	2018-11-01 00:17:28	207	25
8056498	8185645	46838	F	1991	2018-10-31 23:59:47	2018-11-01 00:13:27	207	73

- 2018 y 2019
- 8 millones 56 mil 498 registros





Metodología

Herramientas

- Github
- Google Colab
- OneDrive
- Google Maps
- Heroku
- Python3
 - Pandas
 - Scikit-Learn
 - Googlemaps

Preprocesamiento

- Preparar datasets para su cruce.
- Sacar coordenadas de colonias de delitos y de estaciones de MiBici.
- Estimación de localización de colonias a partir de las coordenadas céntricas.



Análisis

Exploratorio



Exploración general del contenido de los datasets y tipos de distribución

Relacional



Búsqueda de relaciones entre la incidencia delictiva y el uso del sistema MiBici

Modelado



Pronóstico de la ocurrencia de delitos específicos, de acuerdo a la zona frecuentada por usuarios de MiBici



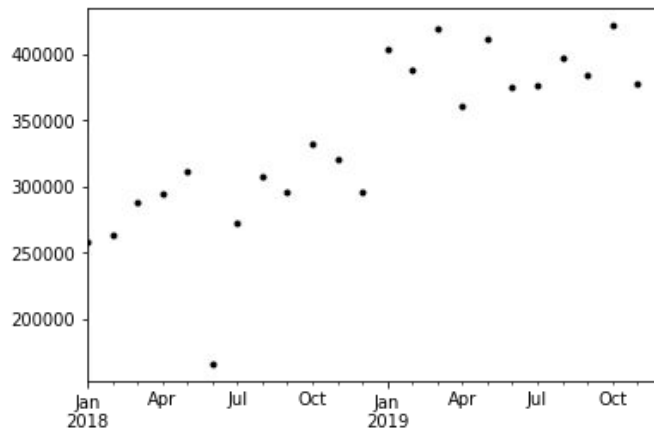
Saturdays.AI



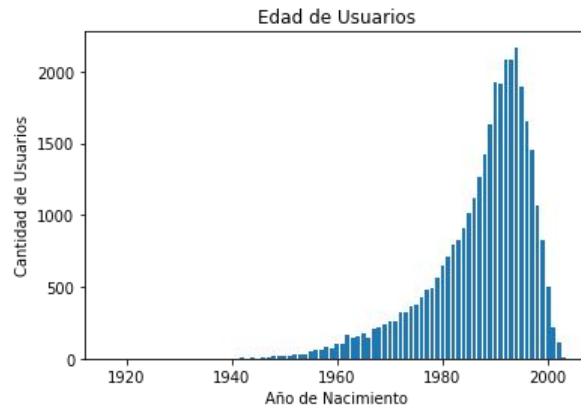
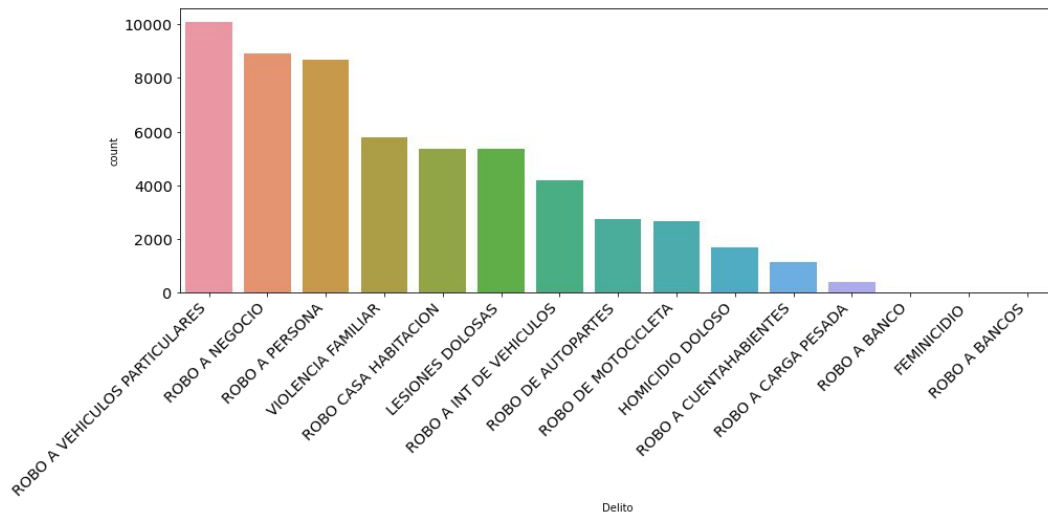
Saturdays.AI
Guadalajara

Análisis Exploratorio

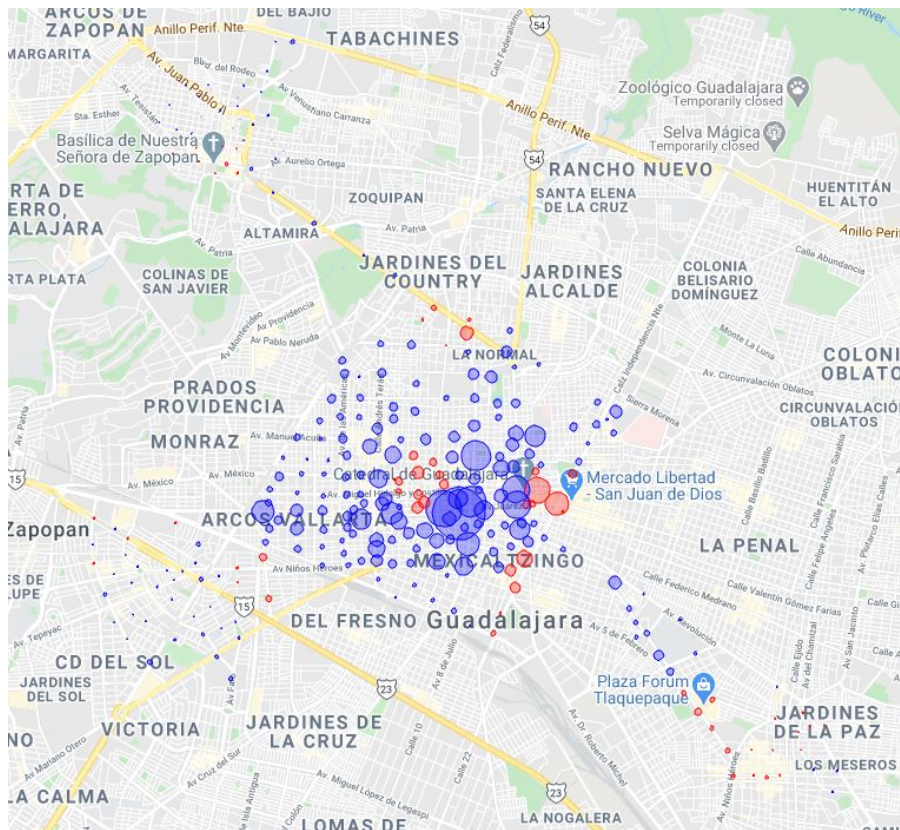
Dataset de MiBici:



Dataset de Incidencia Delictiva:



Análisis Relacional



Saturdays.AI



Saturdays.AI
Guadalajara

Análisis Modelado

¿Qué buscamos?

Pronosticar si un usuario de MiBici puede estar expuesto a un delito en específico.



Saturdays.AI



Saturdays.AI
Guadalajara



Análisis Modelado

¿Cómo lo hicimos?

- Multi-clase
- Entrenamiento y prueba con dataset de delitos.
- Algoritmos: Decision Trees
- Parámetros: Default (pruebas iniciales)
- Tamaño: 16,825 registros (después de filtros)
- Max-Min Scaler

	Número de delitos
Lesiones Dolosas	4480
Robo Motocicleta	2277
Robo a Cuenta-Habientes	1019
Homicidio Doloso	1393
Robo a Negocio	7632
Feminicidio	24



Análisis Modelado

¿Cómo lo hicimos?

Datos de Entrada

Nombre	Descripción
anio	Año en el que ocurrió el delito
mes	Mes en el que ocurrió el delito
location_lat	Latitud de la coordenada de la ubicación del delito. La ubicación consiste en la colonia donde ocurrió el delito ya que es la única información proporcionada por la fuente de datos. Esta información se obtuvo usando el API de Google Maps. Las coordenadas inválidas (fuera de la Zona Metropolitana de Guadalajara) fueron eliminadas.
location_lng	Longitud de la coordenada de la ubicación del delito. La ubicación consiste en la colonia donde ocurrió el delito ya que es la única información proporcionada por la fuente de datos. Esta información se obtuvo usando el API de Google Maps. Las coordenadas inválidas (fuera de la Zona Metropolitana de Guadalajara) fueron eliminadas.

Datos de Salida

Nombre	Descripción
y_lesionesDolosas	1: Si ocurrió delito de lesiones dolosas. 0: Otro tipo de delitos.
y_roboMotocicleta	1: Si ocurrió delito de robo de motocicleta. 0: Otro tipo de delitos.
y_roboCuentahabientes	1: Si ocurrió delito de robo de cuenta-habiente. 0: Otro tipo de delitos.
y_homicidioDoloso	1: Si ocurrió delito de homicidio doloso. 0: Otro tipo de delitos.
y_roboNegocio	1: Si ocurrió delito de robo a negocio. 0: Otro tipo de delitos.
y_feminicidio	1: Si ocurrió delito de feminicidio. 0: Otro tipo de delitos.



Saturdays.AI



Saturdays.AI
Guadalajara

Análisis Modelado

¿Qué obtuvimos? 6 Variables, 16,825 Records

Output	Metric	Valid	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
y_lesionesDolosas	precision	0.25	0.140975	0.12406	0.144893	0.130982	0.145995	0.14
y_lesionesDolosas	precision	0.2	0.115646	0.152429	0.136667	0.148377	0.122481	0.14
y_lesionesDolosas	precision	0.1	0.115242	0.121324	0.127208	0.125	0.138686	0.13
y_roboMotocicleta	precision	0.25	0.062305	0.030374	0.085165	0.062147	0.060942	0.06
y_roboMotocicleta	precision	0.2	0.075949	0.081633	0.04797	0.054054	0.05597	0.06
y_roboMotocicleta	precision	0.1	0.078261	0.052632	0.066667	0.025641	0.061644	0.06
y_roboCuentahabie	precision	0.25	0.019355	0.022388	0.020408	0.017045	0.036585	0.02
y_roboCuentahabie	precision	0.2	0	0.01626	0.016393	0.04878	0	0.02
y_roboCuentahabie	precision	0.1	0.038462	0	0.020408	0.02439	0.037736	0.02
y_homicidioDoloso	precision	0.25	0.055085	0.064327	0.066079	0.066038	0.053763	0.06
y_homicidioDoloso	precision	0.2	0.059211	0.075862	0.071429	0.032468	0.048951	0.06

Análisis Modelado

¿Qué obtuvimos?

2 Variables, 2,412
(Homocidio Doloso / Robo Cuenta-Habientes)

Output	Metric	Valid	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
y_homicidioDoloso	precision	0.25	0.64	0.67	0.66	0.68	0.66	0.66
y_homicidioDoloso	precision	0.1	0.64	0.69	0.65	0.65	0.66	0.66
y_homicidioDoloso	f1	0.1	0.64	0.66	0.63	0.63	0.65	0.64
y_homicidioDoloso	recall	0.2	0.59	0.59	0.66	0.65	0.66	0.63
y_homicidioDoloso	precision	0.2	0.65	0.58	0.57	0.63	0.67	0.62
y_homicidioDoloso	recall	0.25	0.64	0.62	0.63	0.60	0.61	0.62
y_homicidioDoloso	f1	0.25	0.64	0.62	0.64	0.60	0.59	0.61
y_homicidioDoloso	f1	0.2	0.58	0.65	0.63	0.58	0.64	0.61
y_homicidioDoloso	roc_auc	0.1	0.62	0.57	0.61	0.63	0.61	0.61



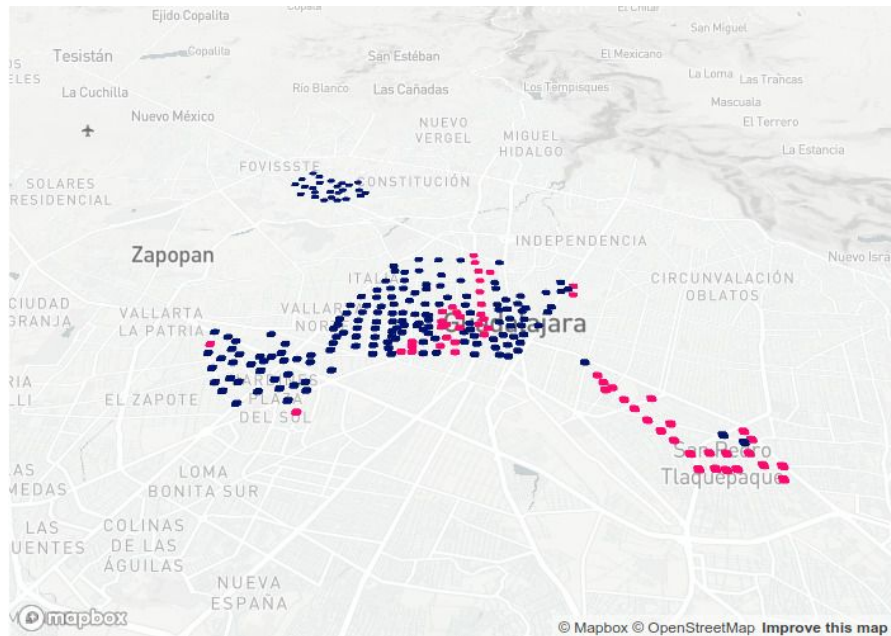
Saturdays.AI



Saturdays.AI
Guadalajara

Análisis Modelado

¿Qué obtuvimos?



Saturdays.AI



Saturdays.AI
Guadalajara



Aprendizajes y Siguietes Pasos

- El **trabajo iterativo** es útil para regresar a fases anteriores cada que se necesite.
- La obtención, **limpieza** y preprocesamiento de datos implica **mucho esfuerzo** y es la base para la **calidad de nuestro análisis** y modelo.
- Un **equipo multidisciplinario** se enriquece de las distintas perspectivas que los compañeros tienen para abordar el mismo problema.
- Para trabajos futuros, se puede explorar un análisis que considere tanto las **variables de tiempo como las variables de espacio** en conjunto.
- Se puede incluir fecha, hora en el set de delitos para mejora del modelo.



Saturdays.AI

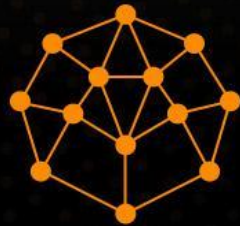


Saturdays.AI
Guadalajara



Recursos

- [Código en repositorio](#)
- [Artículo](#)
- [Documentación del CheckList](#)
- [Mapa con visualización de delitos](#)
- Sitio con visualización de predicciones (link local)



Sataturdays.AI



Sataturdays.AI
Guadalajara

