

PROYECTO FINAL - PRIMER ENTREGA

JUAN SEBASTIÁN ALVAREZ ERAZO
RAFAEL CAMILO TEJON ROJAS
OSCAR JAVIER ÁNGEL BALCÁZAR

CIENCIA DE DATOS APLICADA

DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE LOS ANDES

TABLA DE CONTENIDO

TABLA DE CONTENIDO	2
ENTENDIMIENTO DEL PROBLEMA	3
IDEACIÓN	3
Arquetipos	3
Ideas de solución	4
Storyboard	5
ALCANCE DEL PROYECTO	5
Misión	5
Objetivos estratégicos	6
Implicaciones éticas y riesgos legales	6
ENFOQUE ANALÍTICO	6
Hipótesis	6
Preguntas	7
ENTENDIMIENTO DE LOS DATOS	7
CONCLUSIONES	7
PRÓXIMOS PASOS	8
Creación de la API	8
ANEXOS	8

ENTENDIMIENTO DEL PROBLEMA

La organización con la que vamos a trabajar es la Policía. El problema que logramos observar es que existe una percepción de inseguridad general en todas las ciudades. Ya no existe un horario del día o una zona en la ciudad donde las personas se sientan seguras, lo que genera una incertidumbre en la población para sus quehaceres diarios. Puntualmente el problema se ataca con data de la ciudad de Medellín, la cual es la segunda ciudad más poblada del país después de Bogotá, para el 2020 la ciudad contaba con 2.533.424 habitantes. La ciudad se divide en 16 comunas y zonas las cuales agrupan las comunas, como lo muestra la página [gubernamental](#).

Las entidades del gobierno como la Policía tienen como objetivo velar por la seguridad de la comunidad en general. Se busca agregar valor con los servicios ofrecidos, por ejemplo: ¿Dónde y en qué horarios deben solicitar apoyo de más policías ante un potencial evento de hurto? o, ¿Es suficiente con el CAI ubicado en X lugar para dar abasto ante las situaciones que se presentan diariamente?

Con la data histórica de hurtos en Medellín se pretende tener un primer acercamiento para responder estas preguntas. La data histórica fue obtenida de la página web de la página web de la [Alcaldía de Medellín](#).

Estos son unos ejemplos de métricas que nos pueden ayudar:

- Porcentaje de cantidad de hurtos por hora del día.
- Zonas donde ocurren la mayoría de hurtos.
- Frecuencia de ocurrencia de hurtos.
- Hurtos denunciados versus hurtos ocurridos.

IDEACIÓN

Arquetipos

Policía patrullero	Detalles demográficos	Comportamientos y acciones	Necesidades y dolores
Nombre: Andrés Villanueva 	<ul style="list-style-type: none">- Tiene 34 años.- Casado y con un hijo.- Tiene estudios de primaria y bachillerato.- No tuvo la oportunidad de estudiar una carrera profesional.- Lleva 10 años en la institución.	<ul style="list-style-type: none">- Se preocupa por la seguridad de su ciudad.- Es un policía servicial, ayuda en la comuna donde está asignado cuando hay eventos.- Se enfoca más en las actividades por fuera de la oficina.	<ul style="list-style-type: none">- Cada día nuevo hay un problema diferente.- Le preocupa ser reasignado o un cambio de lugar de trabajo, esto provocaría un cambio de vida para su esposa e hijo.

Directivo de la Policía	Detalles demográficos	Comportamientos y acciones	Necesidades y dolores
<p>Nombre: Juan Torres</p> 	<ul style="list-style-type: none"> - Tiene 55 años. - Está casado y tiene una hija. - Pertenece a la institución desde hace 35 años. - Ocupa un cargo directivo. 	<ul style="list-style-type: none"> - Es una persona que ha mezclado su pasión por la parte administrativa y el servicio a la comunidad. - Ya no ejerce actividades de un policía en la calle, sus labores son más gerenciales y está a cargo de una área importante dentro de la institución. 	<ul style="list-style-type: none"> - Le preocupa la cantidad de nuevos ingresos en la Policía, cada día son menos. - Las herramientas con las que cuenta actualmente no le permiten identificar la distribución de policías por zona. - Siente que durante el día no tiene disponibilidad para atender todos los casos de hurto y violencia que ocurren a diario.

Ciudadana	Detalles demográficos	Comportamientos y acciones	Necesidades y dolores
<p>Nombre: Dalia Gomez</p> 	<ul style="list-style-type: none"> - Tiene 35 años. - Vive en un apartamento con sus 2 hijas que son menores de edad, está separada. - Estudió administración de empresas y está graduada hace 15 años. - Tiene trabajos esporádicos. 	<ul style="list-style-type: none"> - Es una mamá que cuida mucho a sus hijas. - Intenta buscar trabajos que sean de medio tiempo o que sean remotos para poder dedicarle el tiempo necesario a sus hijas. - La separación hizo que se volviera mucho más independiente. 	<ul style="list-style-type: none"> - Siempre intenta buscar un espacio seguro para ella y sus hijas que son su mayor prioridad. - Se moviliza en moto y varias veces ha visto robos. - Inclusive, ha sido víctima de robos por ladrones en moto.

Ideas de solución

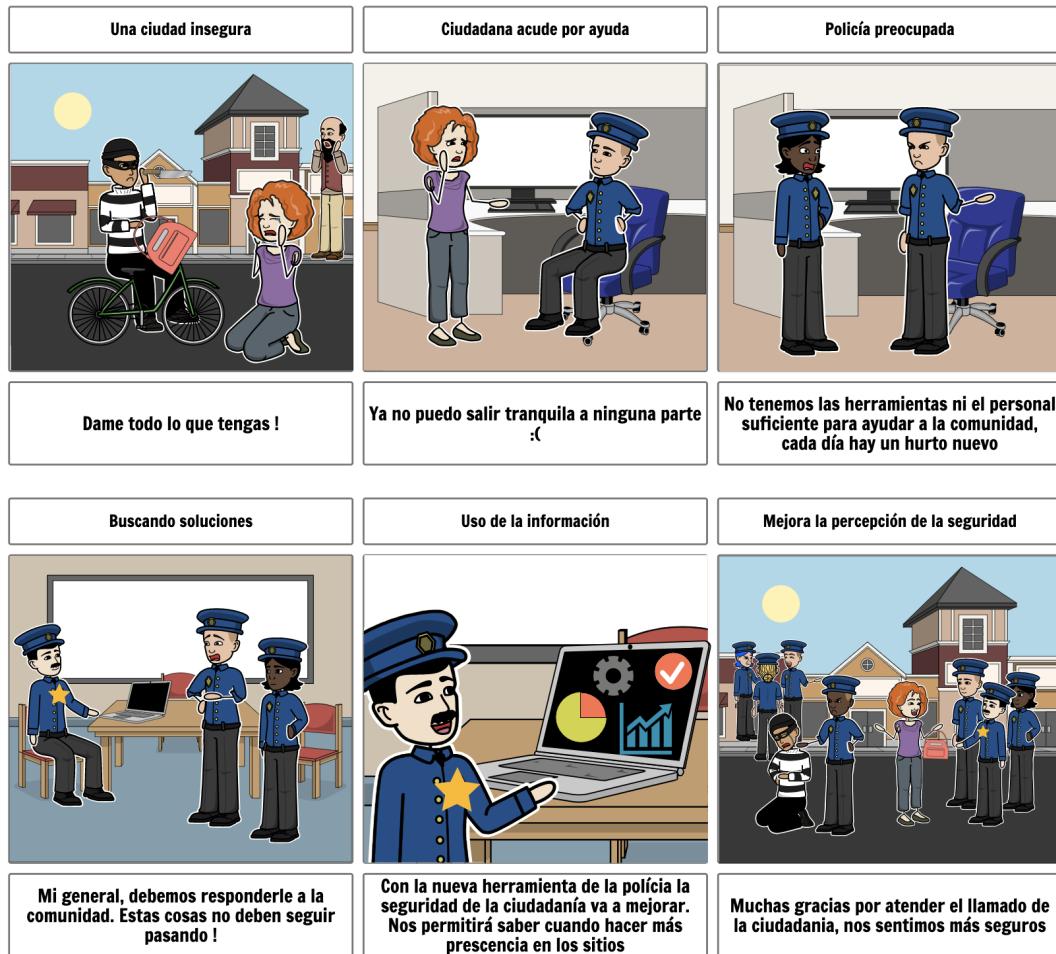
Se plantearon las siguientes ideas para dar solución al problema:

1. Contratar más policías y que puedan ayudar más en la seguridad de la ciudad: No la vemos viable, está fuera del alcance de nuestras posibilidades.
2. Atender más rápido los casos de robo: La policía ya está atendiendo otros asuntos y enfocar todos sus esfuerzos solo en los hurtos no es viable.
3. Construcción de un software con un modelo predictivo: Con la data de hurtos de Medellín se puede crear un modelo y con ayuda de herramientas de Machine Learning hacer que el modelo aprenda y entregue predicciones de los hurtos en una ciudad.

La solución ganadora es la 3. Con la solución ganadora se cumplirán estos objetivos:

- Mejorar la percepción de seguridad de las personas en la ciudad de Medellín de acuerdo a la información de los hurtos.
- Predecir la cantidad de hurtos o hechos delictivos de una zona en unas fechas dadas. Esto permitirá usar mejor los recursos de la Policía.
- Clasificación de zonas seguras y zonas no seguras de una ciudad.

Storyboard



ALCANCE DEL PROYECTO

Misión

Proveer servicios a la policía para complementar la prestación de seguridad y vigilancia.

Visión

Disponibilizar servicios de aprendizaje automático y servicios de predicción que permitan, dadas unas características y entradas, establecer lugares seguros e inseguros en las ciudades.

Objetivos estratégicos

- Aprender de información histórica sobre hurtos en las ciudades y hacer predicciones sobre lo que pasará en el futuro.
- Integrar diferentes servicios de seguridad que proveen información complementaria a los servicios disponibilizados por la compañía.
- Entregar información confiable y construcción de una plataforma escalable, que se ajuste a la demanda del mercado.
- Proteger la información personal de las personas y evitar que esta llegue a manos de personas que la usen con fines diferentes a nuestra misión y visión. Seguimos a cabalidad lo dispuesto en la Ley 1266 de 2008, disposiciones generales del hábeas data.
- Permitir a las personas conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en nuestras bases de datos, nos acogemos a lo dispuesto en la ley de protección de datos personales o Ley 1581 de 2012.

Implicaciones éticas y riesgos legales

- Segregación de comunidades: Pueden ser comunidades que han sufrido desplazamientos, víctimas de la violencia, del conflicto armado, etc. Identificarlos ahora en sus lugares de residencia como “barrios peligrosos” o “sectores no recomendables para vivir” puede generar una revictimización.
- Monitoreo por entes del estado: Por ley nacional la policía, fiscalía, medicina legal y demás entidades del estado tienen autoridad para manejar los datos de hurtos y de las personas y parte de sus funciones misionales es tener el monitoreo de esa información. Siguen la ley del hábeas data y la ley de protección de datos personales. Hay riesgos legales si no se cumplen con las leyes Colombianas.
- Fuentes de información: Siempre se debe obtener información de fuentes de alta confiabilidad para garantizar que los resultados corresponden con la realidad. Obtener información de fuentes no certificadas podría acarrear problemas con los resultados y problemas a las personas que son objeto del estudio. Los datasets obtenidos para esta entrega son de libre acceso y se obtuvieron de páginas gubernamentales: [Portal de datos abiertos MinTic](#) y [Alcaldía de Medellín](#).

ENFOQUE ANALÍTICO

Hipótesis

1. Las mayores víctimas en robos son las mujeres.
2. La mayoría de hurtos ocurren a personas que iban caminando.
3. Las zonas donde ocurren más hurtos están ubicados en el centro de la ciudad.
4. El bien más robado es el celular.

5. Viernes y sábado en horas de la noche son los momentos más peligrosos y donde se presenta el mayor hurto de celulares.

Preguntas

1. ¿Existe una correlación entre la cantidad de hurtos y la cantidad de personas en los sectores de la ciudad?
2. ¿Cuáles son los barrios/comunas más peligrosas de Medellín en los cuales se requiere mayor presencia policial?
3. ¿Cuáles son los tipos de armas más utilizados a la hora de realizar un hurto?
4. ¿En qué se relaciona la edad de la persona y el bien robado/afectado?
5. ¿Cómo ha cambiado la cantidad de hurtos de las diferentes comunas en el tiempo?

Las respuestas a las hipótesis y a las preguntas se encuentran en el [notebook adjunto en el repositorio](#).

ENTENDIMIENTO DE LOS DATOS

Los datos fueron consultados del portal de datos de la ciudad de Medellín, dichos datos se encuentran en formato CSV, el dataset posee 280870 registros con información detallada de los hurtos de la ciudad entre los años 2013 y 2022 con datos como el género, la edad, lugar de los hechos, barrio entre otros.

Dentro del proceso de limpieza de datos identificamos que los nombres de los barrios no estaban tipificados de forma uniforme, por lo cual fue necesario realizar un *lowercase* a varias columnas. También fueron eliminadas columnas como: *cantidad*, *parentesco*, *nivel_academico*, *testigo* entre otras las cuales tenían los valores nulos o constantes en todas las filas del dataset. Con la funcionalidad *PandasProfile* se realizó el análisis y la limpieza de datos. El detalle del anterior proceso se encuentra dentro del [archivo de Colab](#).

CONCLUSIONES

1. Se observa que la ciudad de Medellín tiene un gran problema de seguridad, ya que un alto porcentaje de hurtos se realizan con armas de fuego.
2. El centro de la ciudad es uno de los lugares más afectados por la inseguridad y donde se presenta en mayor porcentaje actividades delictivas.
3. La construcción del modelo predictivo sería ideal que sea general para todas las ciudades, lo que si debe cambiar es la fuente de datos (Dataset, archivos planos, DB, etc.) ya que no podemos predecir hurtos en, por ejemplo, Bogotá con la data de la ciudad de Medellín.
4. Para el modelo predictivo se tiene pensado utilizar una regresión lineal que permitirá predecir la cantidad de hurtos en una fecha dada. Quedará a criterio de la Policía determinar la cantidad de recursos policiales a utilizar en la zona.

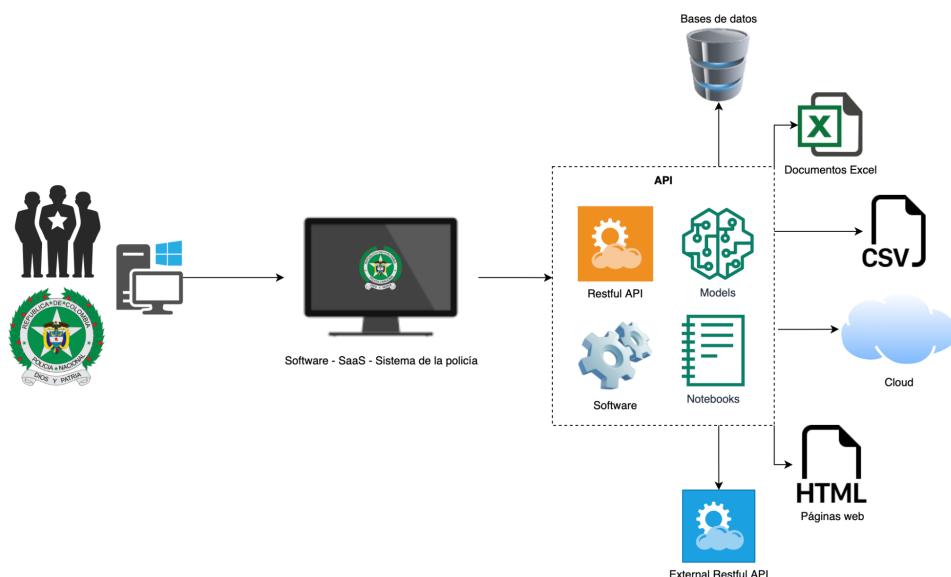
5. La distribución de los robos fue uniforme todos los días, no solo los viernes y sábados donde se pensaba que eran en su mayoría. Ya no depende del día de la semana si te roban o no, sólo puedes sentirte más tranquilo los domingos.

PRÓXIMOS PASOS

Creación de la API

La aplicación será usada por miembros de la Policía. Puede ser una aplicación frontend nueva o una integración en un software existente. Se tiene planeado exponer una API Restful que contiene los modelos analíticos y las herramientas software necesarias para trabajar con datos.

El API se alimentará de información de muchas fuentes de datos como: Bases de datos, Documentos Excel, archivos CSV, software en la nube, páginas web, API externas, etc.



Documentación:

<https://app.swaggerhub.com/apis-docs/JUANSEBASTIANALVAREZ/Policia/1.0.0>

La API Inicialmente contará con 2 end-points:

- Uno que permitirá la consulta de hurtos en la ciudad, se puede filtrar por fechas.
- Y otro que permite conocer la predicción de hurtos en una ciudad en la semana siguiente o en el mes siguiente. Entregará información de los lugares donde podrían suceder estos hechos.

ANEXOS

- [Repositorio](#)
- [Notebook](#)