Reporte Escrito



Pablo Velázquez A. Profesor: Dr José Emilio Quiroz Asesor: Dr Nelson Muriel

La inseguridad en la CDMX: El robo

Prólogo

El proyecto final de ciencia de datos es, entre otras cosas, una recopilación de lo aprendido en el curso y en general de todo lo que hemos visto a lo largo de la carrera. El reporte escrito y la presentación final es simplemente el producto de todo el esfuerzo que se ha llevado a cabo durante un semestre muy complicado debido a la pandemia. Por eso me gustaría empezar por reconocer el esfuerzo de mis compañeros, maestros y de todo el departamento por seguir adelante a pesar de los contratiempos. Igualmente, me gustaría hacer una mención especial a mi maestro de ciencia de datos y a mi asesor. El trabajo que voy a presentar a continuación no hubiera sido posible sin su guía y liderazgo. Sin más por decir, espero que les guste el trabajo que van a leer a continuación y la presentación precedente.

Planteamiento del Problema

Latinoamérica, lugar donde geográficamente está posicionado México, es la región más violenta del mundo. Según un reporte de la BBC "El continente americano reúne 37% de los homicidios de todo el planeta" a pesar de que "concentra apenas el 8% de la población mundial". Algunos de los factores mencionados son la pobreza, la fuerte presencia del crimen organizado, "la guerra contra las drogas", el acceso a armas de fuego, el endeble estado de derecho y la desigualdad. Todos presentes, de alguna manera u otra, en México. Los homicidios no son la única forma de violencia e inseguridad pero son un buen indicador de que tan bien o mal van las cosas en una región o país. Que México esté en latinoamerica no significa que automáticamente sea violento y que el problema está creciendo, pero los números lo afirman. La tasa de homicidios va a la alza y el número total de crímenes por año también está teniendo un incremento considerable en los últimos años.

Como todos nuestros presidentes del siglo XXI saben, la violencia e inseguridad es un rompecabezas muy difícil de resolver. Muchos investigadores de diferentes áreas de estudio pasan su vida estudiando la violencia e inseguridad. Han habido grandes avances y descubrimientos como la correlación entre la pobreza y el crimen o el efecto que tiene la impunidad en el incremento de tasas delictivas. Un elemento común en muchas investigaciones es que buscan entender de mejor manera una pequeña parte del rompecabezas. De igual forma, yo no voy a analizar todo el crimen en México si no que me voy a enfocar en un problema en específico. El problema que voy a analizar es el robo en la Ciudad de México.

La razón de estudiar la Ciudad de México se debe a la disponibilidad de la información y a que es en ella donde se encuentra la Universidad Iberoamericana y he vivido desde mi nacimiento. Por otra parte, sólo considere analizar el robo ya que es uno de los principales tipos de delitos que afectan la percepción de seguridad o inseguridad de los ciudadanos y son los que afectan a la mayoría de la población. El principal motivo del robo es una ganancia económica. En contraparte, delitos como violencia familiar, violación u homicido tienen diferentes motivos menos. Asimismo, hay delitos de bajo impacto que no afectan tanto la percepción de inseguridad en las personas como orinar en la vía pública o fraude fiscal. La exclusión intencional de delitos que no son de tipo robo nos va a permitir, en teoría, encontrar patrones de comportamiento u observaciones de relevancia.

Objetivo

El principal objetivo del proyecto es entender cómo funciona el robo en la Ciudad de México. Para lograr ese objetivo vamos a analizar los datos desde diferentes puntos de vista. El análisis va ir de lo general a lo específico. Todo con el propósito de pintar una imagen clara del comportamiento del robo. Una vez que tengamos la imagen vamos a buscar huellas de patrones de comportamiento visibles tanto en los números como en las gráficas. Esperamos encontrar patrones de temporalidad del robo. Esto basado en artículos académicos que demuestran diferentes tipos de comportamientos que adoptan los ladrones. Ya hay para muchas ciudades, como un artículo que habla del sur de California, pero voy a hacer ese análisis para la CDMX. Esperamos que haya discrepancias entre las variables que afectan el crimen en otras ciudades y las variables que afectan el crimen en México.

Justificación

Al principio del curso nos dieron la misión de desarrollar un proyecto de ciencia de datos de un tema que nos interesara. A pesar de dudar un poco al principio, decidí trabajar en un tema que nos afecta a diario: la violencia e inseguridad en México. Desde que era un niño, escuchaba en las noticias y comidas familiares sobre el problema de violencia e inseguridad que se vive en el país. Uno esperaría que con el paso de los años y el cambio de administraciones, el problema fuera decreciendo pero para mala fortuna de todos los mexicanos, el problema sigue siendo vigente hoy en día. Por eso sentí que hacer un proyecto de investigación referente a la violencia e inseguridad era una oportunidad inigualable para comprender el problema que enfrentamos. Igualmente, es un tema que me apasiona mucho porque me gusta creer que sí tiene solución a pesar de que el futuro se ve turbio. Una de las formas de esclarecer el futuro es mirar hacia el pasado. Ver los diferentes factores que han hecho de la paz un sueño más que una realidad. Si entendemos mejor el problema, somos

más capaces de solucionarlo. La Ciudad de México con menos violencia e inseguridad nos beneficia a todos y todas.

Metodología de la investigación

La base de datos con la que estamos trabajando proviene del gobierno de la Ciudad de México y en particular de la Procuraduría General de Justicia de la Ciudad de México. La información es pública. Se sabe que hay un subregistro de delitos reportados ya que mucha gente que sufre un crimen no presentó una denuncia formal. Por eso, el número real de robos es mayor al reportado en los resultados del proyecto. Todo el análisis se hizo bajo el supuesto de que el subregistro de delitos es constante en el tiempo y con relación al lugar de los hechos. En otras palabras, todas las variables tienen una tasa de subregistro similar. En el año 2016 la base presentó un incremento sustancial en todos los tipos de robos, lo cual se puede explicar por un cambio en la metodología de construcción de los datos, por lo que se eliminaron todas las carpetas de investigación anteriores al 2016. Todos los demás cambios a la base de datos original están documentados en el código que se encuentra en GitHub. El código está diseñado para que cualquier persona que lo corra tenga los datos actualizados del mismo día. Dicho lo anterior, hay una línea de código que descarga una base de datos limpia como un csv. Una vez descargada como csv, esa base de datos se vuelve estática. La base de datos usada en las gráficas de Tableau es un csv estático que corresponde a finales de noviembre del 2020.

Una vez establecido el objetivo era importante establecer metas para seguir adelante. Las metas fueron las siguientes:

- 1. Leer sobre el tema de nuestro proyecto
- 2. Descargar la base de datos
- 3. Limpiar la base de datos
- 4. Saber las preguntas que teníamos que contestar con el uso de la base de datos
- 5. Responder las mismas preguntas
- 6. Verificar que las respuestas de las preguntas tenían información útil
- 7. Explorar las respuestas en busca de nueva información
- 8. Exportar la información importante de Python a un csv
- 9. Graficar el csv con las variables de interés en Tableau
- 10. Corroborar los datos de Tableau con los del código
- 11. Adquirir más información de las gráficas
- 12. Dejar sólo las gráficas de interés
- 13. Leer más sobre el tema

Cada meta que se llevó a cabo tuvo su nivel de dificultad. A pesar de eso hubo unas metas más difíciles que otras. En particular fue complicado establecer las preguntas

correctas para hacer. Teniendo tanta información disponible uno se puede quedar paralizado sin saber qué hacer. En mi caso ante todas las preguntas posibles, no supe decidir de inicio, que preguntas hacerme. Después de hacer muchas pruebas, decidí el tipo de información que quería de regreso para poder explicar el comportamiento del robo de mejor manera. También fue difícil escoger las gráficas de interés. Finalmente la decisión estuvo influenciada por mi opinión personal pero también por una suposición del tipo de contenido que los asistentes querían ver.

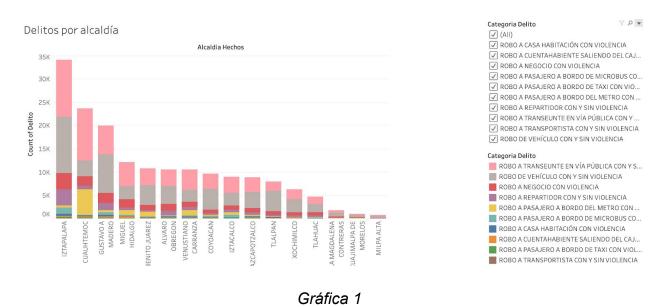
Como pasa en todos los proyectos, incluso aquellos con un objetivo bien establecido desde el principio y una lista puntual de metas a seguir, hubo problemas en el camino. El tipo de problemas a los que me enfrenté varió de problemas técnicos como que el código no funcionara a problemas más fundamentales del proyecto. Tuve que ajustar mi objetivo muchas veces a lo largo del semestre y por consecuencia también mis metas. El primer ajuste que tuve que hacer, y que mencioné brevemente en el párrafo anterior, fue el de escoger las preguntas. Al principio tenía el objetivo de presentar absolutamente todo sobre las carpetas de investigación. Pronto, me dí cuenta de que eso no era posible. La cantidad de información era muy grande y las preguntas demasiadas como para responder todas. Aún más importante, no hubiera sido posible presentar toda la información de forma coherente. Otro problema evidente fue la falta de recursos para demostrar una causalidad. Cuando las gráficas ya estaban hechas y los datos corroborados, fue posible un análisis. El análisis, que era el objetivo final del proyecto, nos dió una muy buena idea de cómo era el robo. A pesar de esto, fue difícil poder establecer causalidad entre variables. Tampoco se pudo demostrar con herramientas matemáticas, como una regresión, si lo que se observaba a simple vista estaba sustentado con números.

El ajuste del objetivo y el ajuste de las metas vino acompañado de un ajuste en la forma de trabajar. Uno de los ajustes que se llevó a cabo en el proyecto fue la incorporación de una API para conseguir la información. Al principio del proyecto, tenía planeado descargar un csv y trabajar directamente con el archivo estático. Este plan cambió cuando mi asesor sugirió que enlazara mi código con la API. A pesar de que fue la decisión correcta a largo plazo, el cambio de planes ocasionó que se dificultará de manera sustancial una meta que en el planteamiento parecía rutinaria. Otro cambio al plan original fue el uso de las diferentes herramientas para desarrollar el proyecto. En un inicio, estaba planeando usar PostgreSQL para el desarrollo del proyecto. Este plan original se vino abajo después de innumerables intentos fallidos de subir la base de datos a la plataforma. Rápidamente Pandas se convirtió en la herramienta predilecta para trabajar. Conforme avanzó el proyecto el plan era continuar trabajando casi exclusivamente con Pandas. En un cambio imprevisto, al final del proyecto dejé de usar Pandas como herramienta principal y cambiamos a Tableau. La razón fue que ver los resultados de manera gráfica permite una mayor compresión en una menor cantidad de

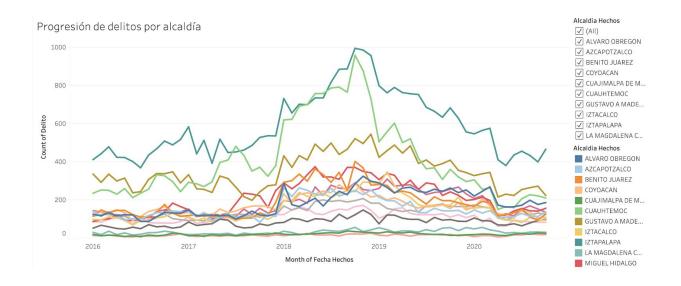
tiempo y hacer más pruebas de forma rápida. Igualmente, Tableau, a pesar de no ser su especialidad, te ofrece la opción de acceder a la información de forma numérica. Otra modificación que le hice al proyecto fue a la hora de presentar los resultados. En un principio se tenía pensado entregar los resultados con una gran carga de elementos matemáticos pero pruebas preliminares arrojaron resultados poco prometedores. La información proporcionada por las diferentes pruebas estadísticas era de poca o nula utilidad. Tomando en cuenta eso, se tomó la decisión de presentar los resultados con un punto de vista más empírico. El cambio a un enfoque empírico ayuda tener una presentación más fluida y menos técnica. Ésto con el objetivo de no saturar a los espectadores con información impractica.

Gráficas descriptivas

Las siguientes gráficas son una captura de pantalla de las gráficas dinámicas hechas en Tableau. El análisis de las gráficas y los resultados están basados en las gráficas de Tableau. Para acceder a las gráficas dinámicas hay un link en la bibliografía.

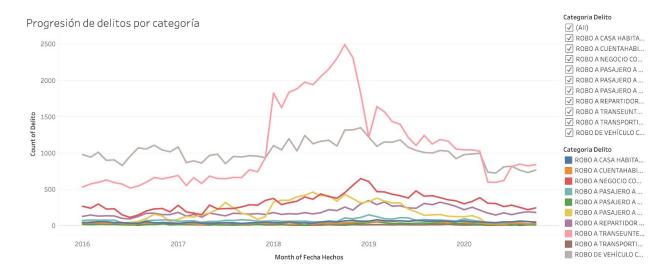


La gráfica 1 es una gráfica de barra que representa el número total de carpetas de investigación por delegación desde el 2016. Está ordenada de mayor a menor. Los distintos colores en las barras se refieren al tipo de categoría de robo.



Gráfica 2

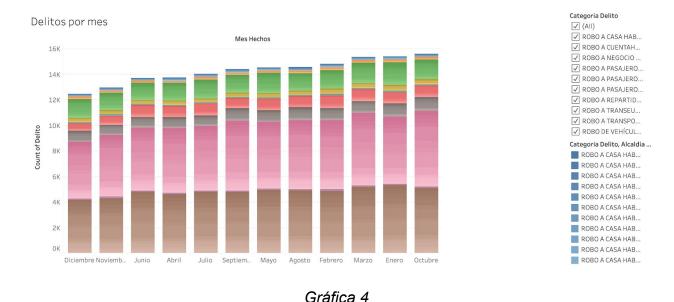
La gráfica 2 es una gráfica de progresión lineal a lo largo del tiempo. En el eje de las x está seleccionada la fecha de los delitos mientras que en el eje y está representado el número de delitos. Cada línea de color representa una alcaldía diferente. Los puntos en las gráficas son referentes a cada mes.



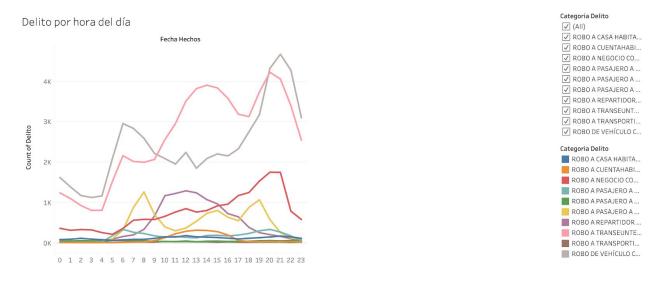
Gráfica 3

La gráfica 3 es muy similar a la gráfica 2. Está gráfica representa una progresión lineal con el paso del tiempo. En el eje de las x está seleccionada la fecha de los delitos

mientras que en el eje y está representado el número de delitos. Cada línea de color representa una categoría de robo diferente. El número de delitos de robo es por mes.



La gráfica 4 es una gráfica de barras que representa el número total de delitos por mes. Al igual que en la gráfica 1, el color de las barras representa el tipo de categoría de robo. La gráfica está ordenada de manera ascendente con la primera barra siendo el mes con menos carpetas de investigación y la última barra siendo el mes con más carpetas de investigación.



Gráfica 5

Está última barra representa la progresión lineal del tipo de robo dependiendo la hora del día. Los números en el eje x representan la hora del día con 0 siendo las 12 de la mañana y 23 siendo las 11 de la noche. Las líneas de colores representan la categoría del delito.

Resultados

El análisis de las gráficas nos da una buena imagen de cómo se comporta el robo en la Ciudad de México. Las gráficas 2 y 3 que son de progresión lineal a través de los años nos demuestran que el número de robos aumentó del 2017 al 2018 considerablemente. El número de robos siguió aumentando hasta finales del 2018 y principios del 2019 donde parece haber alcanzado su punto más alto de los últimos cinco años. Desde principios de 2019 las carpetas de investigación han ido a la baja. También se puede apreciar una baja en los casos de robo en el segundo trimestre del 2020 que corresponden a los inicios de la pandemia. Con el paso de los meses y la apertura de la economía los robos han incrementado nuevamente y mantienen una tendencia creciente.

La primera gráfica nos habla del número de robos por alcaldía. Podemos notar a primera vista la diferencia considerable en el número de robos. Iztapalapa claramente se destaca como el mayor lugar de incidencia siendo la única delegación que rebasa las treinta mil carpetas. Después la alcaldía Cuahutemoc tiene más de 20 mil carpetas y la delegación Gustavo A. Madero está muy cerca de llegar a los 20 mil. Por su parte, Cuajimalpa, Magdalena Contreras y Milpa Alta tienen muy pocos casos reportados. Todas las demás alcaldías se encuentran con números similares. Iztapalapa es frontera con municipios del Estado de México y es una de las zonas más pobres de la Ciudad de México. De igual forma, Gustavo A. Madero colinda con el Estado de México y parte de los municipios más pobres del Valle de México como lo es Ecatepec. Finalmente, Cuauhtémoc tiene dentro de él al barrio de Tepito, zona famosa por su nivel delictivo. Aparte de los factores ya mencionados, la diferencia en número de casos también puede ser un reflejo de la población y de la cantidad de flujo que hay en las calles. Entre más gente haya en la calle, más fácil es escoger el objetivo. Por eso vemos a las delegaciones con una menor población tener un menor número de casos. Finalmente, no es posible apreciar una especialización del tipo de robo por alcaldía. En general todas las alcaldías tienen como sus principales categorías de robo a robo a transeúnte en vía pública y robo de vehículo. Una categoría de robo que si tiene una cierta especialización es la de robo a pasajero a bordo de metro. Esta tendencia tiene que ver con el área geográfica del metro. Mientras en unas delegaciones hay muchas estaciones en otras alcaldías no hay metro por lo que un delito de esta categoría sería imposible. Un ejemplo claro es Cuauhtémoc que tiene un gran número de carpetas de

investigación de robo a pasajero de metro y coincide con que está en el centro de la ciudad.

La segunda gráfica igual corresponde al comportamiento del crimen por alcaldía. Como se planteó en la gráfica 1, los niveles más altos de robo son Iztapalapa y Cuauhtémoc. También se puede apreciar el incremento previamente mencionado en el número de casos a partir del 2018 y la disminución alrededor del 2019. Lo que es interesante, y problemática, de esta gráfica es el aumento a finales del 2017 y la repentina caída del número de casos a principios del 2019. Tomemos por ejemplo a Cuauhtémoc. En noviembre del 2017 se reportaron 322 casos pero para enero del 2018 ese número se disparó a 617 casos. En octubre del 2018 se reportaron 956 carpetas de investigación por robo, tres meses después, en enero del 2019 las carpetas de investigación por robo bajaron a 502. La tendencia continúa con diferentes alcaldías como Benito Juárez, Tláhuac y Xochimilco. Una posibilidad es que a finales del 2017 se hayan relajado los protocolos de seguridad para prevenir robos y un año más hayan incrementado esfuerzos por el alto número de casos, pero otra posibilidad es que el número de casos no tenga nada que ver con el trabajo de las autoridades en la prevención del robo. Puede ser que haya sido una decisión administrativa el redoblar esfuerzos en el registro de las investigaciones y un año más tarde echarse para atrás.

La gráfica tres complementa de manera correcta presentada en el análisis de la gráfica uno y la gráfica dos. Se puede confirmar que los tipos de robo más frecuentes son robo a transeúnte y robo a vehículo. Igualmente refleja la tendencia en el número de casos reportados vista en la gráfica dos. La disminución es notable en el caso de transeúnte que pasa de 738 casos en noviembre del 2017 a 1822 casos en enero del 2018. Después, conforme a la tendencia de la gráfica anterior, el número de robos pasa de 2489 casos en el mes de octubre del 2018 a 1215 casos en enero del 2019. Robo a negocio con violencia muestra un comportamiento similar al de transeúnte. Igual podemos ver cambios causados por la pandemia. Las categorías de robo a pasajero de metro, robo a pasajero de microbús y robo a pasajero de taxi cayeron abruptamente en el número de casos. También las categorías de robo a vehículo y transeúnte se vieron afectadas. La razón es bastante evidente ya que la pandemia causó que hubiera menos gente en la calle o desplazándose de un lugar a otro. En el último trimestre se está viendo un repunte en el número de robos correspondiente a la reapertura de la economía. Curiosamente, el número de robos a casa habitación se ha mantenido constante durante la pandemia a pesar de que la gente pasa más tiempo dentro de sus casas.

La gráfica 4 nos muestra el número de robos por mes. Los dos meses más bajos son diciembre y noviembre pero esto se debe a que todavía no hay datos correspondientes al 2020. Aun así tomé la decisión de dejarlos para tener casi un año más de muestras.

El objetivo de esta gráfica era ver si había un tipo de temporalidad en los robos. Si subía o decrecía el número de robos por estación a lo largo del año o si había categorías más prominentes en ciertas épocas del año. Esto se basaba en conocimiento previo de estudios que demuestran cómo diferentes factores hacen que el crimen varíe a través del año. Lamentablemente, no fue posible encontrar algo evidente. Los tres meses con mayor reporte de robos son octubre, enero y marzo. Todos correspondientes a diferentes épocas del año. Incluso analizando cada categoría de robo de manera individual pero no hay ningún indicio que muestre un cambio en el comportamiento a través del año. Probablemente esto se debe a que a diferencia de otros lugares dónde se han llevado a cabo este tipo de análisis, en la Ciudad de México la temperatura no alcanza niveles tan extremos.

La gráfica 5 muestra el comportamiento del crimen correspondiente a la hora del día. En ella podemos ver como va cambiando el tipo de robo con el paso de las horas. Como era de esperarse, las horas con menos robos es en la madrugada ya que la mayoría de la gente, incluso los ladrones están dormidos. A partir de las 4 empiezan ciertos tipos de robo pero no es hasta las seis donde aumenta esa cantidad considerablemente. El robo a vehículo muestra un claro ejemplo de la temporalidad. Las horas con más número de robos son en la mañana y en la noche. El robo a pasajero de microbús y metro también muestra un comportamiento similar. Una categoría que no comparte este comportamiento es la de robo a negocio. Está actividad delictiva parece ser una actividad vespertina con el número más alto de casos entre las 6 de la tarde y las 9 de la noche. Las categorías de robo a casa habitación, robo a cuentahabiente saliendo del cajero y robo a repartidor son más comunes a la mitad del día. Robo a transeúnte es prominente desde las 11 del día hasta las 10 de la noche. Estas horas corresponden a cuando más gente se encuentra en la calle.

Pronosticar un futuro para el robo parece ser incierto ya que hay una gran variabilidad en el número de casos reportados y estamos en medio de una pandemia que afecta de manera directa el comportamiento de las personas y por consecuencia el de los ladrones. Aún así podemos ver tendencias que nos permitan hacer una suposición fundamentada del futuro. En general, por la tendencia del último año, esperamos que continúe bajando el número de carpetas de robo. Desde el 2019 había una tendencia decreciente y la pandemia sólo lo acentuó. A pesar de que bajen los robos a nivel general, no hay razones para creer que va haber un cambio en el comportamiento de las alcaldías. Las alcaldías que se encontraban hasta arriba van a permanecer ahí y las que se encontraban hasta abajo también. Sólo Cuauhtemoc se espera que retome el segundo puesto con la reapertura de la economía y por consiguiente, el incremento en el número de pasajeros en el metro. El comportamiento de categoría de robo tampoco se espera que cambie mucho, las principales categorías van a seguir siendo las líderes en casos reportados aunque se espera un pequeño incremento en los robos de

transporte público. En el número de delitos por mes puede haber grandes cambios para el 2021. Por lo que sabemos hasta ahorita, el primer trimestre va ser muy similar al 2020 pero si llega a haber una vacuna efectiva y se declara el fin de la pandemia entonces podemos ver un incremento en los siguientes meses. Finalmente, para la hora en la que se cometen delitos predeterminados no se espera que haya algún cambio. Los humanos somos seres muy rutinarios y no se espera que haya alteraciones en el comportamiento que afecten la categoría de robos por hora.

Conclusiones

El trabajo cumple el objetivo planteado al principio de hacer una representación del comportamiento de la violencia e inseguridad, y en este caso del robo en la Ciudad de México. El plan original de presentar absolutamente todo se desvaneció pero eso permitió un análisis de un nicho muy importante del crimen. El desarrollo del nicho da resultados prometedores para seguir trabajando aunque tal vez se pasó por alto datos interesantes por trabajar con una base tan específica. El análisis específico de otras variables quedó como promesa de campaña. La falta de variables hizo que no se pudieran usar otras herramientas matemáticas que nos hubieran proporcionado información interesante. En general, el proyecto presenta buenas bases del entendimiento de lo visto en la clase de ciencia de datos, uso de todas las herramientas proporcionadas y un objetivo claro y bien desarrollado.

En el futuro me encuentro entusiasmado por seguir trabajando con este tema. Siento que hay muchas cosas que se pueden hacer y como dije al principio la violencia y la inseguridad es un rompecabezas muy grande. Entrando más a detalle en mi proyecto, me gustaría pasar más tiempo en la exploración de datos. El análisis de este proyecto es bueno pero es muy general. Quiero ver que otros factores tienen un efecto en el crimen y poder presentarlos, no solo de manera empírica pero también con un sustento matemático. Finalmente, me gustaría hacer de este proyecto un proyecto multitudinario e incluir a gente que complemente mi conocimiento y mis habilidades. Todavía hay mucho por trabajar pero también hemos recorrido un buen tramo del camino.

Bibliografía:

Hipp, John R., and Young-An Kim. "Explaining the Temporal and Spatial Dimensions of Robbery: Differences across Measures of the Physical and Social Environment." *Journal of Criminal Justice*, Pergamon, 6 Nov. 2018, www.sciencedirect.com/science/article/abs/pii/S0047235218303775.

Lissardy, Gerardo. "Por Qué América Latina Es La Región Más Violenta Del Mundo (y Qué Lecciones Puede Tomar De La Historia De Europa)." *BBC News Mundo*, BBC, 12 July 2019,

www.bbc.com/mundo/noticias-america-latina-48960255.

Melgar, Lucía. "Impunidad." *El Economista*, El Economista, 4 Oct. 2017, www.eleconomista.com.mx/arteseideas/Impunidad-20170403-0074.html.

Piccato, Pablo, and Andrés Lajous. "Tendencias Históricas Del Crimen En México." *Nexos*, 1 Apr. 2018, www.nexos.com.mx/?p=36958.

Schuppe, Jon. "Does the Cold Stop Crime? It Seems So." *NBCNews.com*, NBCUniversal News Group, 17 Aug. 2016, www.nbcnews.com/news/us-news/does-cold-stop-crime-it-seems-so-n309856.

"Story Map Journal." *Arcgis.com*, www.arcgis.com/apps/MapJournal/index.html?appid=5508484140a84023a1e2d8b 080e14d0a.

Liga a Tableau

https://public.tableau.com/views/Book1_16057109722410/Sheet4?:language=en&:display count=y&publish=yes&:origin=viz share link