

**Nombre:** David Seijas Pérez  
**URL:** <https://js-wmqucr.stackblitz.io>  
**GitHub:** <https://github.com/davidSeijas/08MBID---Visualizacion-de-Datos>  
**Google Colab:** <https://colab.research.google.com/drive/1WMaMxKoP73fYaegVGdxG7mfjE5T10aa1?usp=sharing>  
**DataSet:** <https://www.kaggle.com/datasets/chicago/chicago-crime>  
**Observaciones:** El dataset se descarga desde los enlaces que hay en la página de Kaggle, que te redirigen a la página de descarga ubicada en la web <https://data.cityofchicago.org/>.

0. ACALARACIONES

Antes de comenzar a exponer cuáles son los objetivos de este trabajo me gustaría hacer una aclaración por la que se guiará todo el desarrollo del trabajo. Este dataset original contenía datos de crímenes en la ciudad de Chicago entre los años 2001 y 2023 (contenía datos actuales). Sin embargo, el volumen de este dataset, contenía más de 7 millones de observaciones, me ha obligado a reducirlo para poder hacer un manejo con algunas herramienta de manera más eficaz y rápida (por ejemplo con Python hubiese podido manejar este volumen, aunque con un aumento del tiempo de respeusta, pero con Google Sheets no y quería tener el mismo dataset para todas las herramientas).

Aún así tanto mi audiencia y mi objetivo están planteados como si hubiese podido realziar gráficos de todos los datos que conteía ese dataset, es decir, como si tuviesémos datos desde el año 2001 hasta el momento actual. Por esta razón, aunque plantea evoluciones entre estos años me gustaría que se considerasen como si fuese hasta la actualidad por lo que conseraré que estamos en el año 2006, ya que la creación de gráficos hubiese sido la misma, simplemente cambia la cantidad de datos.

Por último, me gustaría añadir que los datos de 2001 debían estar mal informados porque hay una clara diferencia entre la cantidad de datos del año 2001 y el resto de años, simplemente he asumido estos datos como buenos porque no influenciaban a la hora de la realización de los gráficos; para que se tenga en cuenta a la hora de la visualziación de los gráficos.

1. INTRODUCCIÓN

Este trabajo consiste en una StoryTelling formado por diversos gráficos sobre los crímenes cometidos en la ciudad de Chicago entre los años 2001 y 2005, es decir, tiene datos de los crímenes cometidos hasta la actualidad.

Este trabajo va dirigido a las autoridades y organismos encargados de la seguridad pública de la ciudad de Chicago. El objetivo de estos gráficos será proporcionar información detallada sobre la evolución de los principales crímenes en Chicago durante los últimos años. Además, destacar las tendencias y patrones específicos de los crímenes, así como los porcentajes de arrestos asociados. Esto podría ayudar a las autoridades a identificar áreas problemáticas y enfocar los esfuerzos en la prevención y el combate de los delitos más frecuentes.

Con estos gráficos y esta presentación se pretende conseguir una mejora en la organización de la seguridad de la ciudad y pdoer centrarse en los aspectos más importantes para pdoer reducir el incide de criminalidad de la ciudad.

Para concluir la presentación, cabe destacar que los gráficos han sido alimentados con datos extraídos de una fuente fiable como es la base de datos "Crimes 2001 to Present" de <https://data.cityofchicago.org/>.

2. METODOLOGÍA

Para alcanzar el objetivo descrito anteriormente se han desarrollado gráficos usando las herramientas "Google Sheets", "Python en Google Colab" y "Qlick Sense". La elección de estas herramientas se ha debido a los siguientes motivos:

- **Google Sheets:** Google Sheets fue la primera herramienta que utilicé debido a su accesibilidad y facilidad para la creación de gráficos sencillos. Además de que es una herramienta sencilla, presenta muchos diseños de gráficos distintos y así pude ubicarme con el dataset y conocer más el objetivo que pretendía alcanzar. Sin embargo, cuando quise utilizar gráficos más complejos tuve que cambiar de herramienta debido a que Google Sheets no está pensada para gráficos complejos. Otro problema que encontré fue la lentitud a la hora de procesar los gráficos debido a que el volumen de mis datos era bastante grande y no los soportaba del todo bien.
- **Python:** Fue la segunda herramienta que utilicé ya que, debido a mi objetivo, quería realizar gráficos más complejos, además de mapas y gráficos interactivos y esta herramienta me era más sencilla de utilizar para dicho propósito. Quería alcazar gráficos más personalizados, además de más complejos, y Python me permitía esto. Por último, sabía que iba a poder procesar y diseñar los gáficos de forma mucho más rápido ya que es una herramienta que soporta un gran volumen de datos. Por suerte, tengo experiencia programando y eso me hice que no me costara adaptarme y crear visualziaciones de manera veloz, pero es cierto que para gente sin experiencia puede ser bastante complicada de usar.
- **Qlick Sense:** Por último, usé la herramienta de Qlick. Lo malo de esta herramienta es que es un poco más compleja ya que ofrece muchas alternativas y, además, es de pago. Pude usarla gratis a una prueba gratuita y tampoco pude profundizar tanto como quisiera en los gráficos ya que es cierto que no era lo más intuitiva. Sin embargo, tenía claro que quería utilziar esta herramienta ya que era un reto nuevo, aún por conocer, y que me permitía el diseño de un dashboard interactivo de una forma seniclla, dentro de lo que cabe.

A continuación, deajo una tabla con los puntos fuertes y débiles de cada herramienta utilizada, con sus respectivas referencias.

Herramienta	Puntos Fuertes	Puntos Débiles
Google Sheets	Es una herramienta muy accesible para usuarios de todos los niveles de habilidad. Su interfaz intuitiva y sus funciones básicas lo hacen fácil de usar (Palma et al., 2021).	A diferencia de otras herramientas de visualización de datos, como Python o Qlik Sense, Google Sheets puede tener limitaciones en la personalización de gráficos. Esto puede hacer que sea difícil crear gráficos altamente personalizados y complejos (Strauss, 2023).
Google Sheets	Herramienta ideal para crear gráficos básicos y simples que se utilizan comúnmente en informes y presentaciones. Puede crear gráficos de	Limitaciones en la cantidad de datos que puede manejar. Si se están procesando grandes conjuntos de datos, es

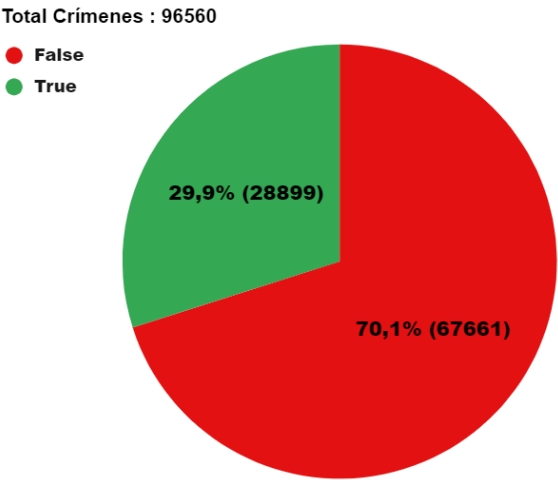
	barras, gráficos circulares, gráficos de líneas, entre otros (Rosa & Rivera, 2016)	posible que se experimente una disminución en la velocidad y el rendimiento (Urbina, 2018)..
Google Sheets	Accesibilidad y almacenamiento en la nube, lo que permite el acceso desde cualquier dispositivo con conexión a internet (López, 2017).	Es menos estable que otras herramientas cuando se trabaja con grandes conjuntos de datos, lo que puede llevar a la pérdida de datos o al tiempo de inactividad (Strauss, 2023).
Python	Variedad de librerías especializadas en gráficos como Matplotlib, Seaborn, Plotly y Bokeh, que permiten la creación de gráficos complejos e interactivos (Koehrsen, 2019).	Tiene una variedad significativa de librerías de gráficos, es posible que no ofrezca tantas opciones como otras herramientas de visualización de datos (Heer & Bostock, 2010).
Python	Capacidad para automatizar el proceso de creación de gráficos, lo que ahorra tiempo y aumenta la eficiencia en el análisis de datos (Koehrsen, 2019).	Python presenta barreras de entrada para aquellos que no tienen experiencia en programación. Necesita de unos conocimientos básicos sobre ello (Rodríguez, 2016).
Python	Flexibilidad en la personalización de gráficos, que permite a los usuarios ajustar aspectos como colores, tipografía, etiquetas y leyendas para adaptarlos a sus necesidades específicas(Koehrsen, 2019).	Es menos eficiente que otras herramientas de visualización de datos en términos de velocidad (Manrique, 2020).
Qlick Sense	Permite la creación de dashboards altamente interactivos e intuitivos que permiten a los usuarios explorar y analizar datos de manera fácil y eficiente (Fernández, 2022).	Es una herramienta de visualización de datos de pago, lo que significa que su costo puede ser una barrera para algunos usuarios. Aunque hay una versión gratuita disponible, la mayoría de las funciones avanzadas solo están disponibles en la versión de pago. (Yépez, 2020).
Qlick Sense	Permite la exportación de dashboards y visualizaciones en una variedad de formatos, incluyendo PDF, Excel y formatos de imagen. Esto significa que los usuarios pueden compartir fácilmente sus análisis y visualizaciones con otras personas dentro y fuera de la organización (Fernández, 2022).	Hay un límite de caracteres para las expresiones que se utilizan en la creación de gráficos. Este límite puede variar dependiendo de la versión de Qlik Sense que se esté utilizando (Yépez, 2020).
Qlick Sense	Ofrece acceso a través de API, lo que permite a los usuarios integrar fácilmente las visualizaciones y dashboard en otras aplicaciones y sistemas. También se puede acceder a través de dispositivos móviles, lo que significa que los usuarios pueden acceder desde sus teléfonos y tabletas en cualquier momento y desde cualquier lugar (Fernández, 2022).	Puede tener una curva de aprendizaje empinada para los usuarios nuevos o principiantes (Fernández, 2022).

3. RESULTADOS

Lo primero que me gustaría plantear a la audiencia es: ¿Cuántos crímenes han sucedido entre el año 2001 y 2005 en Chicago y qué porcentaje de esos han acabado con el delincuente arrestado?

Imagen 1. Desarrollado en Google Sheets

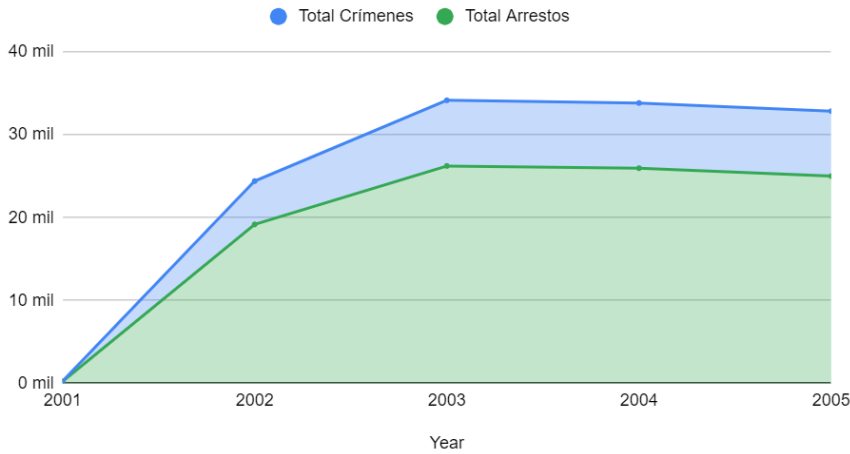
Porcentaje de Arrestos en Crímenes



Viendo estos datos demoledores, en los que de cada 10 crímenes cometidos en la ciudad más de 7 salen con sus delincuentes impolutos, podríamos preguntarnos si en los últimos años se ha producido un aumento o bajada de los crímenes y, además, de los arrestos producidos, para ver si hemos obtenido alguna mejora con el desarrollo de los últimos años. La respuesta es NO.

Imagen 2. Desarrollado en Google Sheets

## Evolución Crímenes y Arrestos con los Años

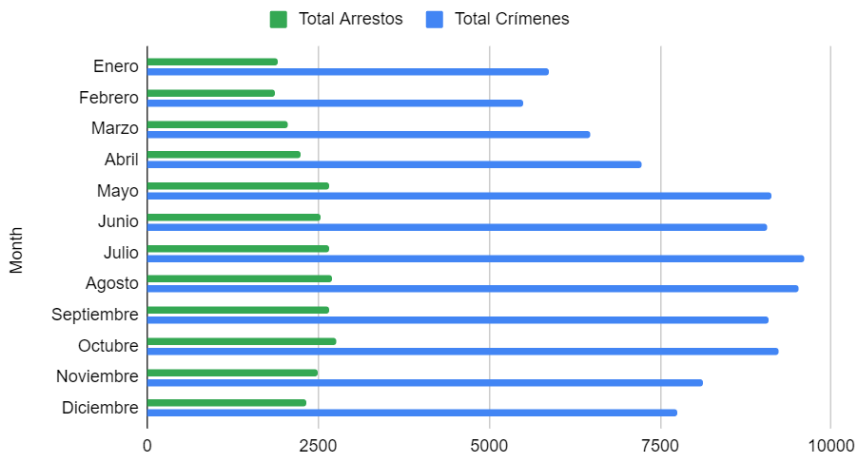


En este segundo gráfico (gráfico de líneas, el mejor gráfico para representar evolución o tendencias) podemos ver la evolución que han tenido tanto los crímenes como los arrestos en estos últimos años. Quitando el caso especial del año 2001, vemos que del año 2002 al 2003 se incrementaron los crímenes, habiendo unos 10 mil más, quedándose casi constante desde entonces.

Ya hemos visto que queda mucho por mejorar en el tema de la seguridad en cuanto a crímenes. Entonces empezamos por estudiar información de años atrás para en qué crímenes debemos enfocarnos y cuánto. Entonces la siguiente pregunta que quiero preguntar es, ¿en qué momento, días, meses, del año suceden más crímenes? Si supiéramos cuando hay más crímenes, podríamos centrarnos en reforzar la seguridad esos meses.

**Imagen 3.** Desarrollado en Google Sheets

## Crímenes por Mes



Lo que podemos observar en este gráfico es la cantidad de crímenes cometidos en cada mes del año desde 2001, además del número de ellos que acabaron en arresto en cada mes. Como podemos ver los meses de verano y próximos es cuando más crímenes se dan, así que deberíamos esforzar en reforzar la seguridad esos meses, puesto que los crímenes crecen considerablemente, pero no los arrestos.

Ahora, no vale solo con aumentar la seguridad, hay que saber dónde focalizarse, por eso la siguiente pregunta que quiero plantear es, ¿cuáles son los 10 crímenes más cometidos a lo largo de estos años? ¿y cuáles son los 10 donde se han hecho más arrestos?

**Imagen 4.** Desarrollado en Google Sheets

## Top 10 Tipos de Crímenes Cometidos

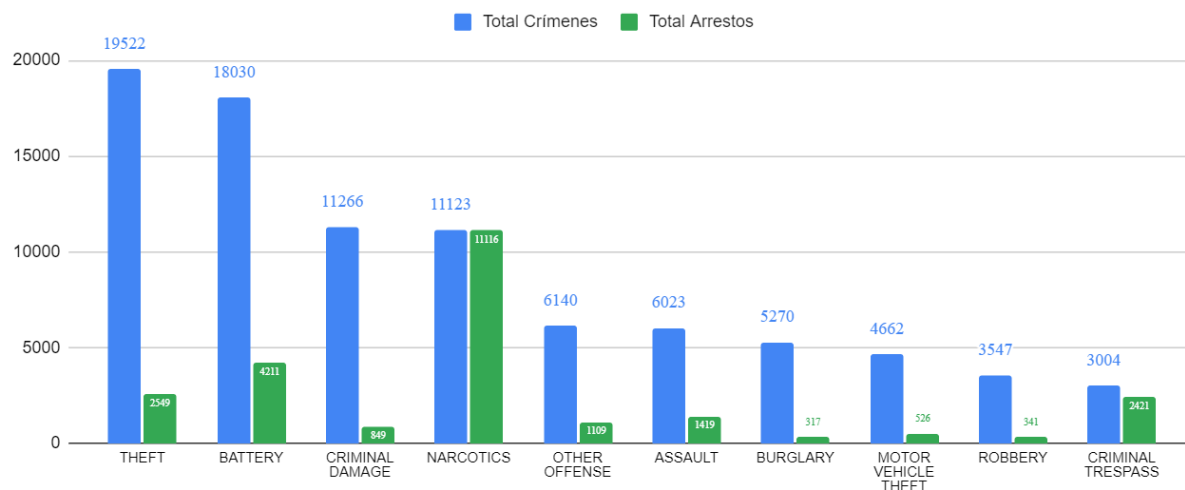
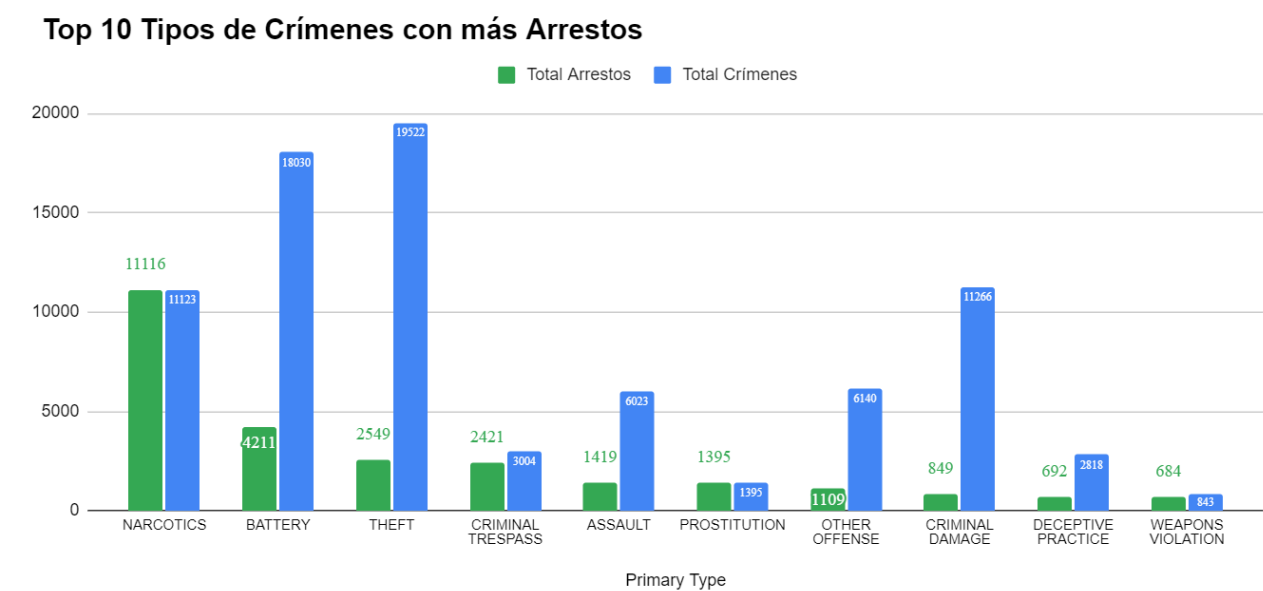


Imagen 5. Desarrollado en Google Sheets

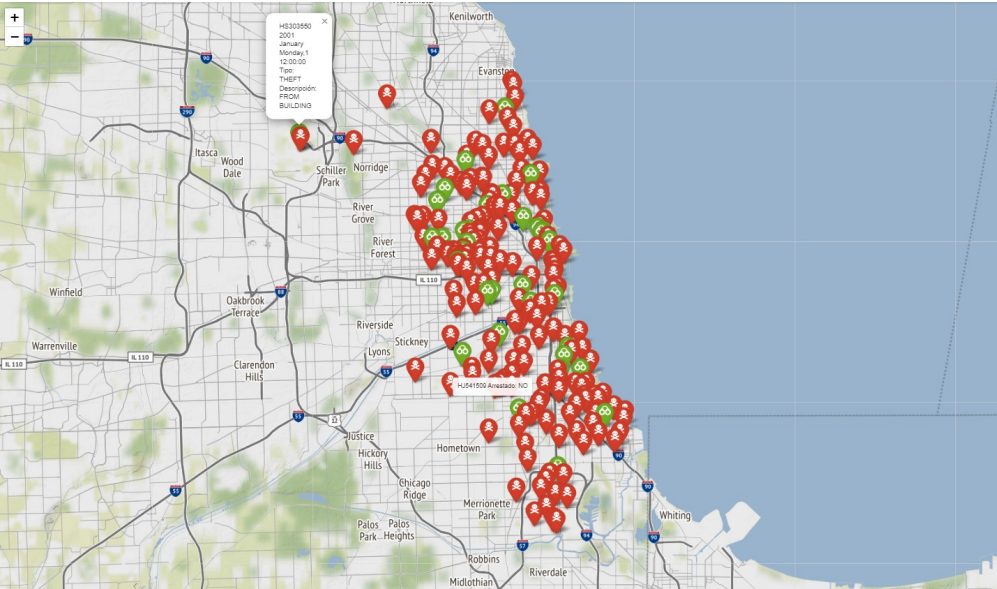


Estos gráficos (de columnas, perfectos para realizar comparaciones) exponen los 10 tipos de crímenes con más casos en estos 5 años y sus respectivos arrestos; y los 10 tipos de crímenes donde se ha arrestado a más delincuentes con sus respectivos número de casos.

Observando estas imágenes podemos llegar a la conclusión que hay que reforzar especialmente los robos y las agresiones o daños criminales, ya que han sido los crímenes más cometidos estos últimos años, pero el número de arrestos sucedidos no llega a un número aceptable. Igual los dos últimos crímenes son más difíciles de predecir y actuar contra ellos, pero hay un grave porblema con los robos que se debe afrontar. Por el otro lado, podemos observar que en el seguimiento de criminales de *narcotráficos* hay un gran desempeño.

Estos datos están muy bien y ya tenemmos una idea de los crímenes en los que hay que prestar más atención y los meses en los que habría que reforzar la seguridad, sin embargo, la ciudad es muy grande y hay que saber dónde. Por ello mi siguiente gráfico responde esa pregunta de ¿dónde han sucedido los crímenes en los últimos 5 años?

Imagen 6. Desarrollado en Python. En el enlace a Google Colab se puede ver el mapa y sus interacciones.



En este mapa podemos ver marcados en la ciudad de Chicago todos los crímenes cometidos en estos 5 años. Los crímenes que acabaron con arresto salen marcados en verde, mientras que los que no están en rojo. Además, si pasas por encima de cada crimen te confirma el número del caso y si acabó en arresto o no, mientras que si pinchamos en un crimen nos saca la información del año, mes, día y hora en los que sucedió.

Ahora que sabemos en qué momento del año hay reforzar la seguridad y sabemos en qué zonas sucedieron más crímenes los últimos 5 años y, por tanto, es más probable que vuelvan a suceder en el futuro, nos deberíamos preguntar en qué momentos del día suceden más crímenes. Y voy más allá, no sirve de nada saber que a las 13:00 suceden más crímenes de narcóticos, pero reforzamos para estar atentos a robos. Por tanto, otra de las preguntas importantes que deberíamos plantearnos es ¿en qué momento del día suceden más crímenes de cada tipo? A estas dos preguntas dan respuesta los siguientes gráficos.

Imagen 7. Desarrollado en Python.

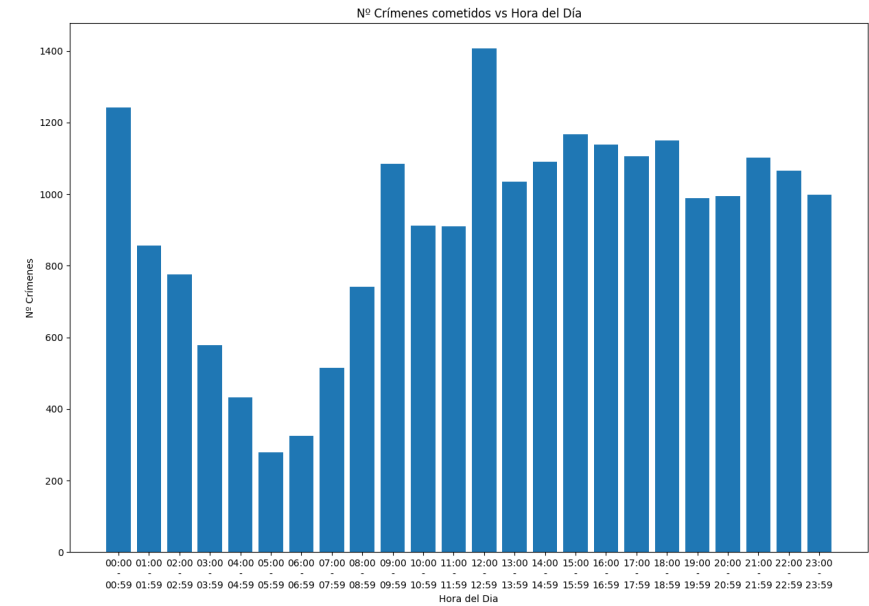
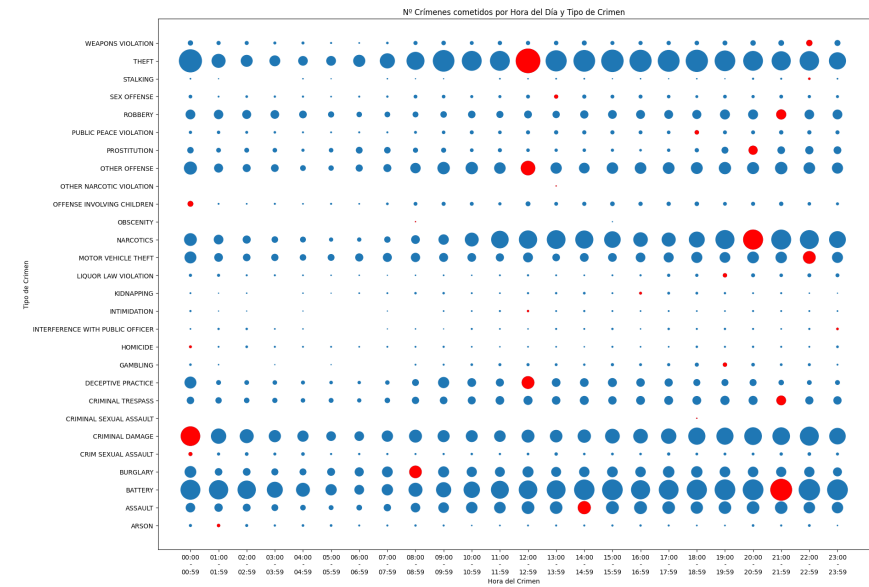


Imagen 8. Desarrollado en Python.



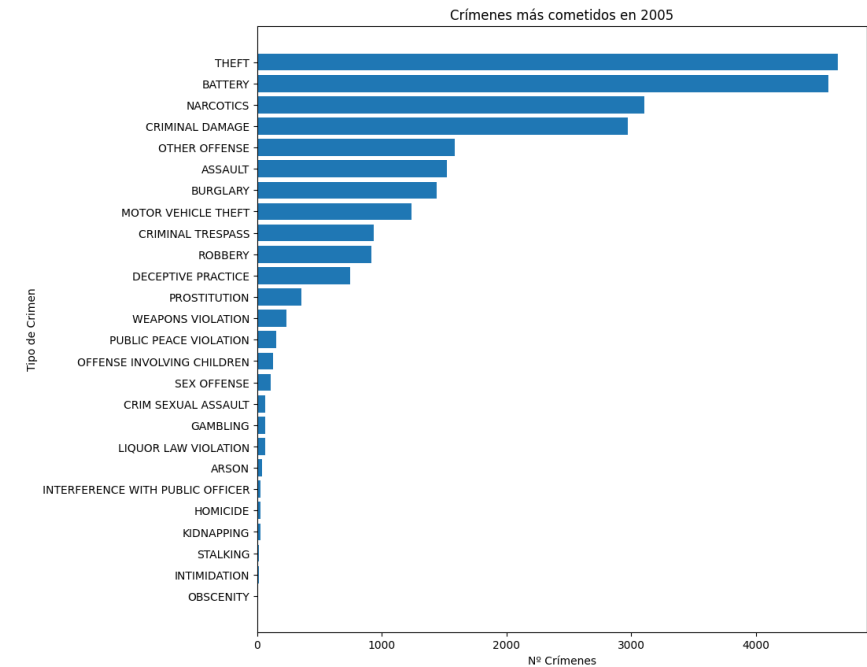
En la Imagen 7 podemos observar como, aunque nos pueda sorprender, la hora del día donde más crímenes se cometieron fue las 12 del mediodía. Sin embargo, como hemos mencionado esta imagen tampoco nos aportando tanto valor, sino que tenemos que fijarnos en la segunda.

En la Imagen 8 podemos ver un gráfico de dispersión, elegido pues es ideal para representar 3 variables y sus relaciones, que representa la cantidad de crímenes cometidos (mayor tamaño de círculo, mayor número de crímenes cometidos) por tipos de crimen y en cada hora del día. Los círculos resaltados de color rojo representan la hora del día en la que más crímenes se cometió de cada uno de los tipos.

De esta manera, si nos fijamos en los robos, recordemos que era el crimen más cometido en los últimos 5 años, cuando más se comete es a las 12 del mediodía, mientras que las agresiones sexuales se cometen más por las noches, en especial a las 12 de la noche.

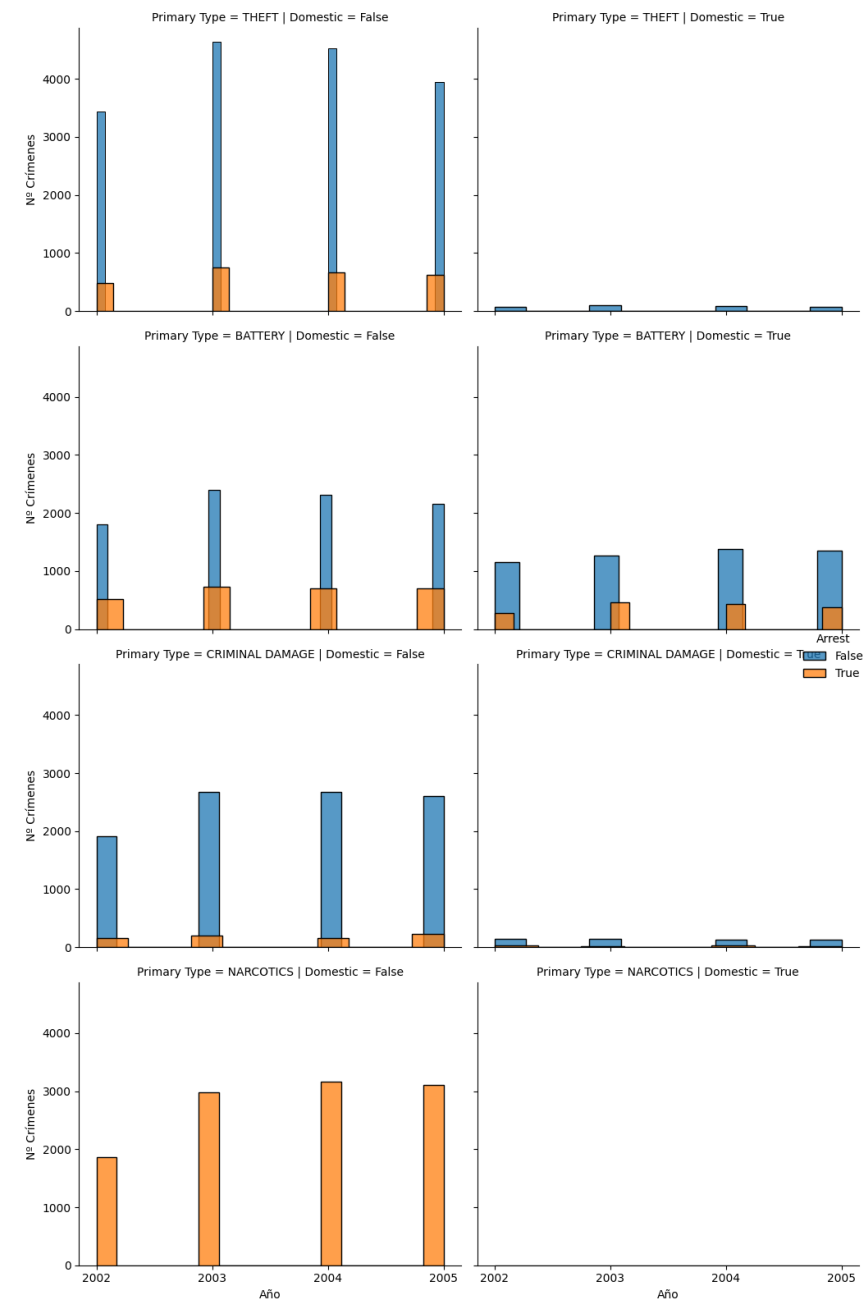
Me gustaría continuar nuestro análisis de los crímenes centrándonos concretamente en el último año, el año 2005. Ya que son los datos más recientes, me gustaría realizar el estudio de cuáles fueron los crímenes más cometidos para ver si la tendencia está cambiando o sigue siendo similar a la global de los últimos años.

Imagen 9. Desarrollado en Python.



En la Imagen 9, gráfico de barras elegido para comparar los crímenes más cometidos en el año 2005, podemos ver se mantienen los mismo crímenes como los más comunes en el último como en la visión global. Visto esto, me gustaría plantear a la audiencia, ¿cómo ha sido el balance de estos crímenes en los últimos 5 años (obviaremos 2001 por falta de datos) y cuántos se han cometido en el ámbito doméstico?.

**Imagen 10.** Desarrollado en Python.

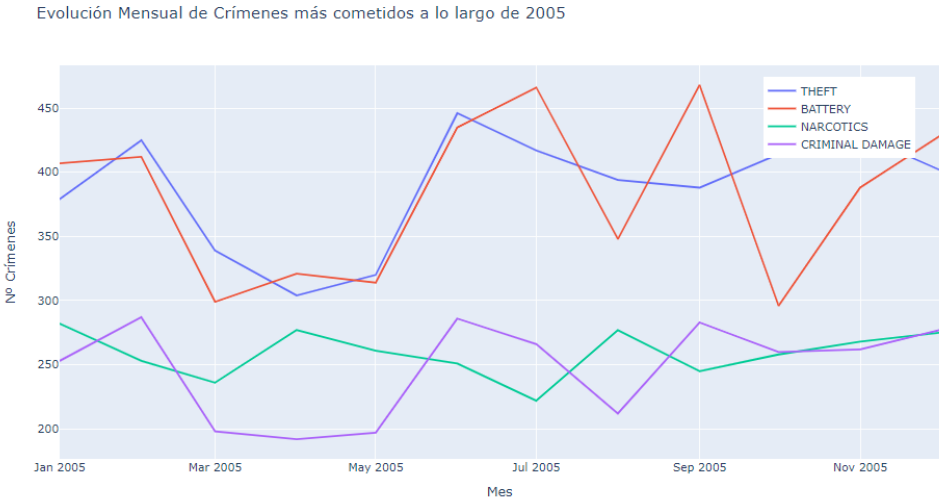


El gráfico correspondiente a la Imagen 10 es una matriz de gráficos que representan cuántos crímenes acabaron en arresto y cuántos no a lo largo de los últimos según el tipo de crimen y si este se realizó en ámbito doméstico o no. Estos gráficos son interesantes a la hora de realizar comparaciones separando entre una o variables variables que consideramos significativas.

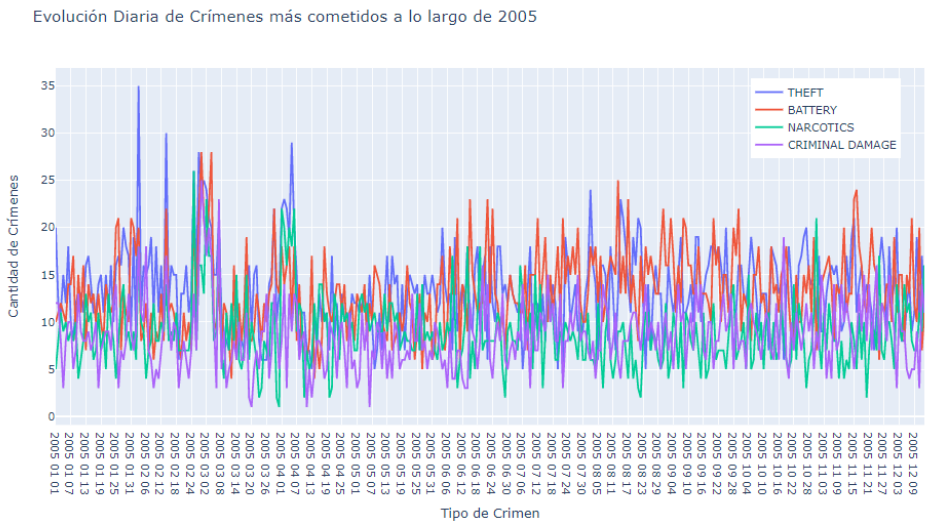
Como entre los 4 tipos más comunes aparecen los robos y agresiones me resulta interesante saber cuántos se producían en un entorno familiar para saber cómo abordar la seguridad y evitar que se comentan tantos crímenes. De hecho, observamos que hay un gran número de agresiones domésticas que habrá que tener en cuenta y considerar para evitar que se realicen.

Para seguir con el análisis de los crímenes del último año, he realizado dos gráficos interactivos que muestran la evolución tanto mensual como diaria de los 4 crímenes más cometidos este año. Con estos gráficos me gustaría responder a cuándo se han producido los picos de más crímenes de cada tipo en el último año para tratar de evitarlos en años venideros.

**Imagen 11.** Desarrollado en Python. Puede verse la interacción en Google Colab.



**Imagen 12.** Desarrollado en Python. Puede verse la interacción en Google Colab.



Aunque a primera vista, sobre todo la Imagen 12, puedan parecer confusos, es un gráfico interactivo que permite hacer zoom y sacar la información de cuántos crímenes sucedieron cada día en concreto, esclareciendo mucho la pregunta que nos hacíamos.

Para concluir el análisis de los crímenes cometidos en la ciudad de Chicagi en los últimos, me he querido centrar en un tipo de crimen que ha sucedido mucho para la gravedad que conlleva y que se ha saldado con muy pocos arrestos, con el problema que esto supone. Este crimen son las agresiones sexuales. Me gustaría, además, centrarme en este delito pues con los años se está normalizando.

Por esta razón, he desarrollado un dashbord que abordea esta problemática. Este dashboard da respuesta a unas preguntas que creo necesaria que se haga la audiencia para poder evitar estos delitos desde ya: ¿cuántos delitos de agresión sexual se han cometido en los últimos años? y ¿cuántos de estos han acabado con arresto del agresor?, ¿cómo han evolucionado las agresiones sexuales desde el 2001?, ¿en qué época del año se cometen más? y ¿en qué momentos del días?

**Imagen 13.** Desarrollado en Qlick.

Nº de Crímenes Sexuales

313

**Imagen 14.** Desarrollado en Qlick.

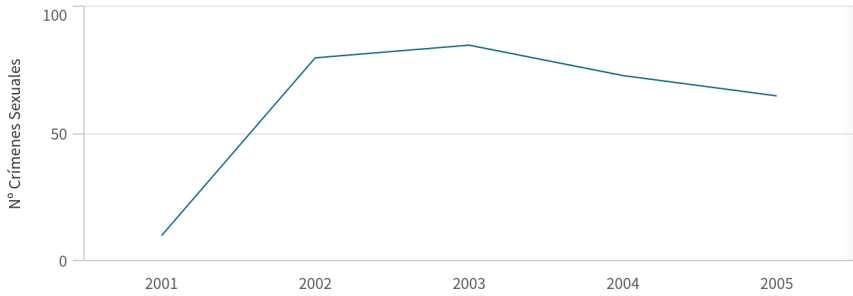
Nº Arrestos Crim. Sexuales

49

**Imagen 15.** Desarrollado en Qlick.



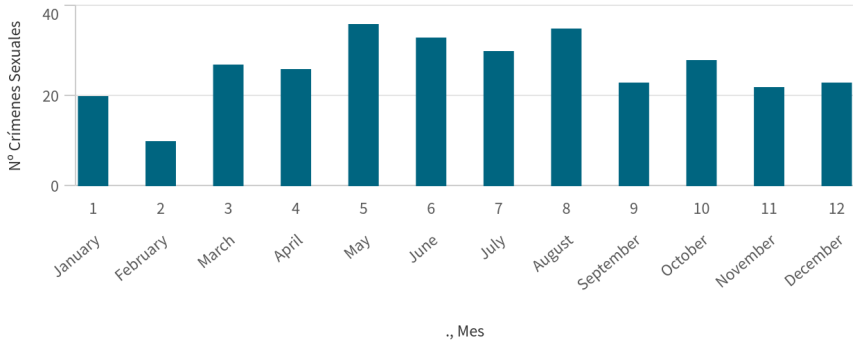
Evolución Crímenes Sexuales



Filtros aplicados: Primary Type: CRIM SEXUAL ASSAULT, CRIMINAL SEXUAL ASSAULT

Imagen 16. Desarrollado en Qlick.

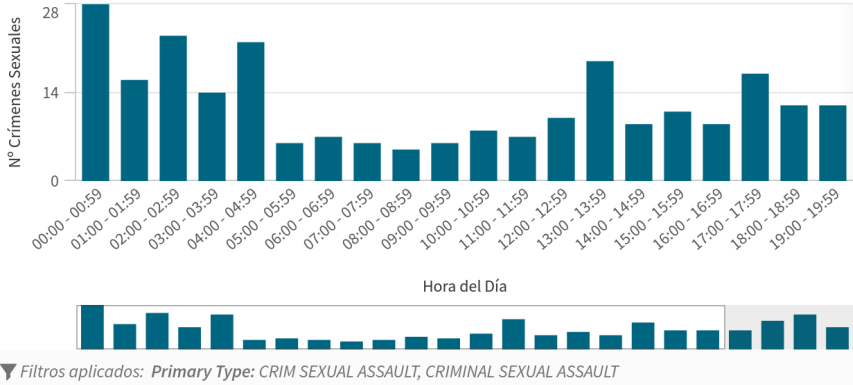
Crímenes Sexuales vs Mes del Año



Filtros aplicados: Primary Type: CRIM SEXUAL ASSAULT, CRIMINAL SEXUAL ASSAULT

Imagen 17. Desarrollado en Qlick.

Crímenes Sexuales vs Hora del Día



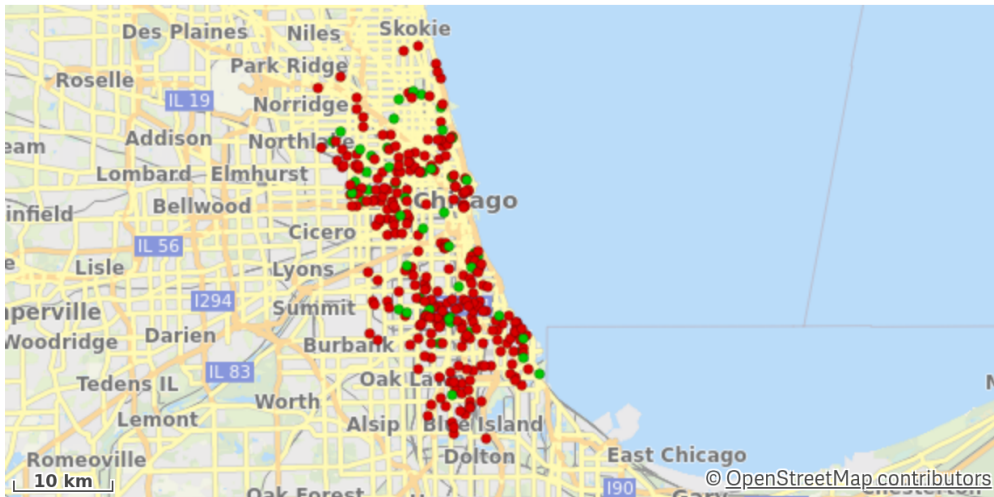
Filtros aplicados: Primary Type: CRIM SEXUAL ASSAULT, CRIMINAL SEXUAL ASSAULT

Observando estas imágenes podemos concluir que esto es un grave problema a resolver. Si nos fijamos en las Imágenes 16 y 17, las agresiones sexuales suceden más en meses de verano y durante las horas de la madrugada. Podemos llegar a la conclusión que en meses donde se producen más fiestas y a las horas donde la gente está en la calle es cuando más se producen estos delitos. Observando estos datos, es claro que hay que reforzar la seguridad estos meses a estas horas de la madrugada.

Sin embargo, ¿dónde se producen estos delitos? Muchas veces estos delitos se suceden en calles oscuras o de poca afluencias. Para dar respuesta a esta cuestión, he realizado un último gráfico que es un mapa donde se han producido los crímenes sexuales (resaltando en verde los que acabaron en arresto y en rojo los que no). Además, como he hecho en el mapa anterior, al pasar por encima de cada punto nos da la información concreta de cada crimen.

Imagen 18. Desarrollado en Qlick.

## Lugar de los Crímenes Sexuales



Este dashboard que he realizado es interactiva y si seleccionamos datos de algunos de los gráficos se filtra en todos los gráficos y muestra la información de lo que estamos seleccionando exclusivamente (se ve el comportamiento en el vídeo adjunto). Además, se puede modificar el filtro para visulizar la información correspondiente a otros tipos de crímenes de forma muy sencilla.

A todas estas preguntas dan respuesta, respectivamente, los siguientes gráficos incluidos en el dashboard.

## 4. CONCLUSIONES

Para concluir con este trabajo me gustaría comentar que yo creo que los objetivos se han cumplido concretos. He descubierto nuevas herramientas, además de aprender nuevos gráficos, entre ellos gráficos interactivos, de herramientas que ya conocía como Python. Además, me he dado cuenta de la potencia de Python como herramienta de visualización a pesar de no ser una herramienta específica para esta tarea. Comentar también que la herramienta de Qlick me ha resultado muy interesante y muy útil. Creo que tiene un potencial enorme tanto para desarrollar gráficos interactivos como para personalizar diseños de dashboards.

He disfrutado muchísimo realizando este trabajo, pensando en qué gráficos realizar e investigando cómo personalizar estas o cómo ir un pasito más allá para que representen lo que quiero decir exactamente. Desarrollando esta memoria me he dado cuenta de la verdadera importancia de fijar el objetivo de las visualizaciones y la audiencia a quien va dirigida.

Por último, me gustaría comentar ciertos aspectos que pueden quedarse como puntos a mejorar en el futuro. El primero, es poder integrar los gráficos de Python y de Qlick en esta memoria final, investigando sé que con Qlick sería posible aunque por falta de tiempo no he podido averiguar cómo y si es necesaria la versión de pago. Además, me hubiese gustado tocar otras herramientas que, por falta de tiempo como PowerBI, o por imposibilidad debido a sus características como DataWrapper, no he llegado a usar en este trabajo. Considero que cuando te pones a realizar gráficos más complejos por tu cuenta es cuando realmente aprendes a usar a fondo una herramienta, que es lo que me ha ocurrido con las que he usado yo.

## 5. REFERENCIAS

- Fernández, O. (2022). *Herramientas BI: Comparando las Mejores Opciones*. Aprender Big Data. <https://aprenderbigdata.com/herramientas-bi/>
- Heer, J., & Bostock, M. (2010). *Crowdsourcing graphical perception: using mechanical turk to assess visualization design*. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203-212. <https://doi.org/10.1145/1753326.1753357>
- Koehrsen, W. (2019). The Next Level of Data Visualization in Python. *Towards Data Science*. <https://towardsdatascience.com/the-next-level-of-data-visualization-in-python-dd6e99039d5e>
- López Toledo, E. (2017). Gestión de Mantenimientos en la nube mediante Google. *Revista Iberoamericana de Producción Académica y Gestión Educativa*, 4(8). <https://www.pag.org.mx/index.php/PAG/article/view/690>
- Manrique Rojas, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E28), 586-599.
- Palma Kuhl, C., Capp, E., & Henrique Nienov, O. (2021). Google apps: Drive, Slides, Forms, Docs e Sheets. *Estrategias didácticas para actividades remotas*, 47-58.
- Rodríguez Ojeda, L. (2016). *Python Programación*. Escuela Superior Politécnica del Litoral.
- Rosa Urrutia, V. I., & Rivera Pleitez, J. G. (2016). Big Data, análisis de datos en la nube. *Entorno*, (61), 17-24. <https://biblioteca2.utec.edu.sv/sitios/entorno/index.php/entorno/article/view/493>
- Strauss, L. (2023). *Google Sheets vs. Excel: Which is right for you?* [2023]. Zapier. <https://zapier.com/blog/google-sheets-vs-excel/>
- Urbina, V. (2018). *Los límites de google sheets: macro para monitorear el tamaño de un archivo*. Fditf. <https://www.fditf.com/bi/google-sheets/los-limites-de-googlesheets-macro-para-monitorear-el-tamano-de-un-archivo/>
- Yépez Palacios, M. B. (2020). *Analítica De Datos Aplicada a la Estimación y Predicción de Ventas para la Empresa Textil Saga de la Ciudad de Atuntaqui Utilizando Qlik Sense* [Tesis Doctoral]. Pontificia Universidad Católica del Ecuador Sede Ibarra.