



LABS PROYECTO GRUPAL

SINIESTRALIDAD VIAL NEW YORK

INTEGRANTES:

**JUAN DIEGO GUTIÉRREZ
JEAN CARLOS BETANCOURT
YEINER ALEXANDER IMBACHI
CARLOS GAVIRIA**

BOOTCAMP HENRY

9/09/2022

CONTENIDO

1.Entendimiento del problema	3
2. Objetivo	5
3. Alcance del Proyecto	5
4. Fuera del Alcance	6
5. Solución Propuesta Stack tecnológico	6
6. Metodología de Trabajo	7
7. Diseño detallado-Entregables	8
8. Equipo de Trabajo- Roles y responsabilidades	8
9.Soluciones del Producto	9
10.Manual de Instrucciones	10
9. Cronograma General	16
Referencias	17

1. Entendimiento del problema

La siniestralidad vial hace referencia al conjunto de eventos trágicos causados generalmente por errores humanos los cuales se podrían haber evitado. En la ciudad de Nueva York, una de las ciudades más grandes del mundo con aproximadamente 8.38 millones de habitantes para el 2020, diariamente se registran cientos de eventos de siniestralidad vial, dada la gran cantidad de medios de transporte. Por tanto, la municipalidad de NYC, requiere encontrar nuevas soluciones que permitan mejorar la calidad de transporte y movimiento dentro de la Ciudad haciéndola más segura.

La división del área de Transporte, registra los eventos de siniestralidad que suceden diariamente en la ciudad, y pone a disposición las bases de datos necesarias para llevar a cabo el análisis, a las cuales se puede acceder mediante API'S y/o exportación en formato CSV. Adicionalmente, dicha información se podrá complementar con fuentes de datos externas

Se analizará el tráfico vehicular y la accidentalidad vial en la ciudad de New York para tratar de establecer las causas mediante la relación de algunos eventos especiales, tales como:

- Flujo vehicular en horas pico
- Flujo vehicular en cada día de la semana
- Nivel de accidentalidad

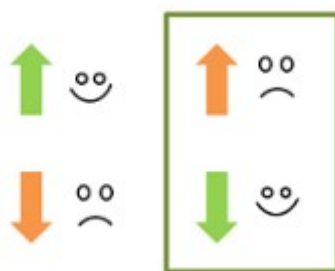


Figura 1. Función de Optimización

La función de optimización que indicaría (en este caso en específico de los siniestros en NYC) que se está cumpliendo con el objetivo, es sí la variación porcentual es negativa (disminuye) es un buen indicador y si aumenta no se está logrando, lo que se desea de forma general es disminuir los siniestros.

Todo esto para establecer las siguientes KPI:

- **Indicador Decremental de Siniestros Graves:**

Siniestros Graves, hace referencia a sólo aquellos sucesos en los que resultan personas fallecidas tomando en cuenta Peatones, Ciclistas y Conductores.

Este KPI indica la variación del número de muertes de forma trimestral. El objetivo es que en cada periodo trimestral, los siniestros graves identificados, se reduzcan en un %5 tomando como referencia el trimestre inmediatamente anterior. Por tanto, siempre se busca que la variación sea negativa y menor al 5% para que se cumpla el objetivo para dicho trimestre.

- **Indicador Decremental de Siniestros Leves:**

Siniestros Leves, hace referencia a sólo aquellos sucesos en los que resultan personas heridas pero no fallecidas, tomando en cuenta Peatones, Ciclistas y Conductores.

Este KPI indica la variación del número de accidentes leves de forma trimestral. El objetivo es que en cada periodo trimestral, dichos siniestros, se reduzcan en un %5 tomando como referencia el trimestre inmediatamente anterior. Por tanto, siempre se busca que la variación sea negativa y menor al 5% para que se cumpla el objetivo para dicho trimestre.

- **Indicador Decremental de Siniestros Relacionados a Ciclistas:**

Los siniestros relacionados con ciclistas incluyen las muertes y heridos ocurridos para dicho grupo. Por tanto, este KPI indica la variación porcentual del número de siniestros con respecto al trimestre anterior, cuyo objetivo es lograr una disminución de, al menos, un 5%. Un porcentaje negativo y menor a dicho porcentaje, indica que se cumplió el objetivo para dicho trimestre..

Adicionalmente, se analiza la relación entre la totalidad de siniestros ocurridos y el factor contribuyente, para identificar el motivo que más siniestros causó en dicho periodo.

- **Indicador decremental de Siniestros Relacionados con Peatones:**

Los siniestros a peatones incluyen personas fallecidas y heridas en dicho grupo. Este KPI tiene objetivo reducir el número de siniestros en un 5% de forma trimestral, donde una variación negativa y menor a 5% con respecto al trimestre anterior indica el cumplimiento de dicho objetivo. Adicionalmente, se analiza la relación entre el número de siniestros y el factor contribuyente, con el objetivo de identificar la causa más común de siniestros para el trimestre en cuestión.

- **Indicador decremental de Siniestros Relacionados con Conductores:**

Los siniestros en conductores incluyen personas fallecidas y heridas. Este KPI tiene objetivo reducir el número de siniestros en un 5% de forma trimestral, donde una variación negativa y menor a 5% con respecto al trimestre anterior indica el cumplimiento de dicho objetivo. Adicionalmente, se analiza la relación entre el número de siniestros y el factor contribuyente, con el objetivo de identificar la causa más común de siniestros para el trimestre en cuestión.

- **Indicador Decremental de siniestros por Carretera:**

En este indicador se tienen en cuenta las muertes y personas heridas por carretera incluyendo peatones, ciclistas y conductores. Muestra una variación porcentual con respecto al trimestre anterior, cuyo objetivo es lograr una reducción del 5%, por tanto, al darse un porcentaje negativo menor, se está cumpliendo con el objetivo para dicho trimestre. También, se analizan las carreteras mayor tráfico y su correspondencia con el total de siniestros para identificar las vías que presentan mayores eventos. Adicionalmente, se tiene en cuenta la información de los factores contribuyentes, para determinar cuáles de ellos son los más comunes para determinada carretera.

2. Objetivo

General:

Proponer alternativas para la movilidad vehicular, que permitan reducir la siniestralidad vial de la ciudad de Nueva York

3. Alcance del Proyecto

- Overview

Se trata de un sistema para el análisis de datos que facilite la proposición de alternativas para el mejoramiento de la calidad de transporte y movilidad, dada la gran problemática de siniestralidad vial y accidentalidad que presenta la Ciudad de Nueva York. Para ello, se usarán bases de datos facilitadas por la División de Transporte de la Ciudad, e igualmente fuentes de información externas que permitan el análisis adecuado. Adicionalmente, se usarán herramientas software para tratamiento y sistemas de servicios en la nube que permiten alojar millones de datos sin necesidad de requerir hardware dedicado. Finalmente, el usuario final podrá visualizar la información para la posterior toma de decisiones.

- Se cuenta con la aprobación de:
Henry

- Se realizará teniendo en cuenta las siguientes fases:

Fase 1: Recolección y almacenamiento de datos en Data Lake desde diferentes fuentes (API, csv).

Fase 2: Limpieza y transformación de datos (ETL) desde el data lake para su posterior análisis, adicionalmente se subirán los datos al data warehouse localizado en la nube y de esta forma se otorgará acceso a los datos para diferentes fines

Fase 3: Análisis de los datos desde el data warehouse y visualización de los mismos a partir de diferentes métricas

Fase 4: Control de calidad y documentación para la respectiva entrega.

- Dentro de lo que se producirá se lista lo siguiente
 - Algoritmo que permita la selección de datos y su posterior transformación
 - Se hará entrega de un Data warehouse
- Se hará entrega de un Dashboard el cual contenga:
 - Un mapa de NYC que muestre las zonas geográficas con mayor concentración de accidentes, calles y avenidas más utilizadas.
- Calculadora que permita predecir siniestralidad para determinada fecha y por distrito.
- Se estima que la fecha de entrega sea:
Viernes 9 de Septiembre, 2022.

4. Fuera del Alcance

Se consideran fuera de alcance todos aquellos procesos adicionales que se agregaron fuera de las fechas estipuladas para el proyecto. De ser necesario un ajuste del alcance, se revisarán los cambios y se analizará si la solicitud es lo suficientemente importante como para ser tomada en cuenta en el proyecto y de esta forma evitar una corrupción en el mismo. Tampoco se considera la formulación del presupuesto del proyecto, costos y gastos.

5. Solución Propuesta Stack tecnológico

Por medio de diferentes herramientas como lo son Python, AWS, PowerBI, Streamlit, se llevará a cabo el proceso del ciclo de vida del dato, desde la adquisición, transformación, almacenamiento, visualización, e inteligencia artificial, la cual permita mostrar información relevante, a partir de la cual se pueda tomar decisiones importantes.

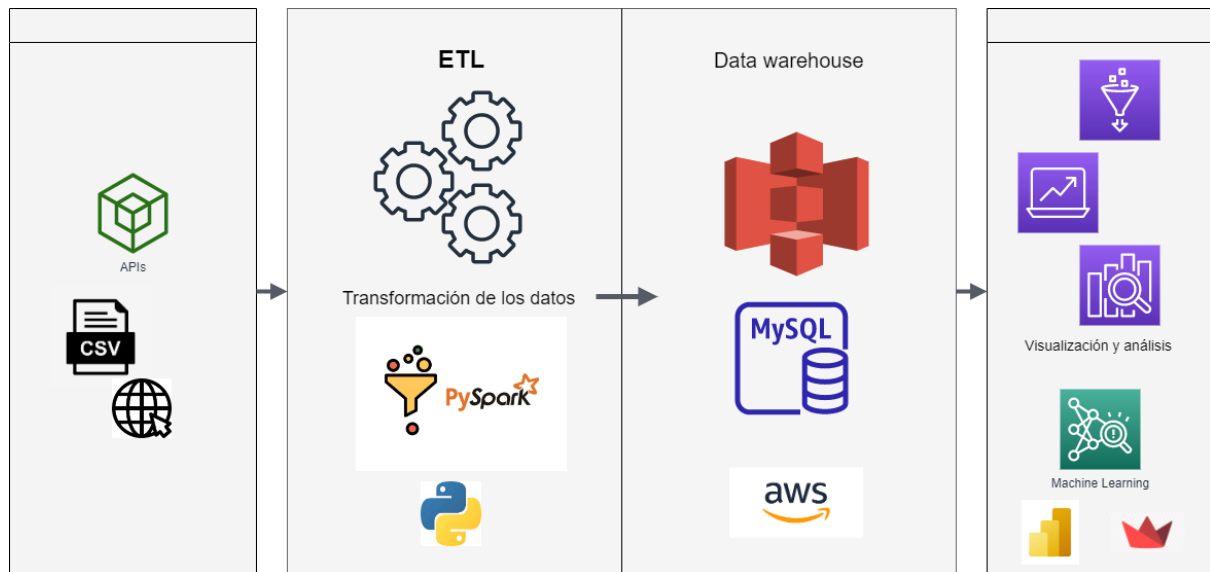


Figura 2. Propuesta Tecnológica

Se usarán dichas herramientas ya que ofrecen diferentes tecnologías:

- Python: Se creará el script transformador que permite limpiar y mejorar el output de los datos, facilitando de esta forma el trabajo de análisis.
- AWS: Este servicio en la nube ofrece diferentes métodos de almacenamiento además posee rápido/fácil acceso a los datos sin comprometer la seguridad.
- PowerBI: Herramienta de visualización perfecta para el análisis, se contará con datos limpios directamente desde el data warehouse para visualizar los diferentes KPIs, métricas, gráficos y mapas para entender mejor la solución al problema.
- Streamlit: Plataforma de visualización para aplicaciones de análisis, se maneja directamente desde el código del programa y permite mejor manejo para visualizaciones de machine learning.

6. Metodología de Trabajo

La metodología SCRUM, hace parte de una de las metodologías ágiles de software, que integra prácticas eficaces para el trabajo colaborativo y en equipo, que se basa en entregas parciales las cuales están jerarquizadas de acuerdo al aporte de cada una al proyecto, y de las cuales se obtiene realimentación por parte de los usuarios finales y sus necesidades[1].

Esta metodología, ofrece cinco grandes ventajas[2]:

- Satisfacción al cliente:
El cliente se encuentra directamente comprometido con el desarrollo y el resultado final del proyecto, pues de él depende qué y cuándo se hace determinada actividad, mientras que el equipo de trabajo le indica con qué se puede comprometer.

- **Simplicidad:**
Cada actividad se encuentra parametrizada, indicando los roles que participan para su desarrollo y el tiempo necesario.
- **Inspección:**
Permite ver qué tan bien se está llevando el proyecto y qué cosas deben mejorar.
- **Adaptación:**
Permite el cambio de características del producto en cualquier momento, siempre y cuando no se afecte el concepto de la entrega pactada.
- **Trabajo en equipo:**
Las personas que forman parte del equipo logran adaptarse al ciclo de desarrollo, a través de la acción conjunta de los participantes, donde cada individuo se considera como parte fundamental, tanto así, que el cambio de uno de los participantes puede generar gran impacto para el producto final.



Figura 3. Herramienta de Gestión de proyectos

Como herramienta de gestión de proyectos, se usará Trello, dado que brinda flexibilidad, permite diseñar planes, organizar flujos de trabajo y hacer un seguimiento del progreso de una manera visual, productiva y gratificante. Trello gestiona las tareas diarias, desde la lluvia de ideas y planificación hasta la ejecución del proyecto[3].

7. Diseño detallado-Entregables

- El sistema software de análisis completo, desde la obtención, transformación, almacenamiento, visualización de la información y predicción de siniestralidad vial
- Documentación

8. Equipo de Trabajo- Roles y responsabilidades

El proyecto contará con la participación de 4 miembros listados a continuación y sus respectivos roles:

- Carlos Gaviria (Scrum Dev, Data Engineer/ML Engineer)
- Juan Diego Gutiérrez (Scrum Dev, Data Engineer)
- Jean Carlos Betancourt (Scrum Dev, Data Analyst)

- Yeiner Imbachi (Scrum Dev, Data Analyst)

En Calidad de Scrum Master:

- Marcos Audisio

En calidad de Product Owner:

- Dario Miguel Lopez

9.Soluciones del Producto

La solución que se da con el producto es brindarle al usuario información clave y detallada en los diferentes distritos de la ciudad de Nueva York indicando y separando por los diferentes involucrados definidos en tres grupos: ciclistas motoristas y peatones. Adicionalmente le permite identificar cuál es el factor contribuyente que más siniestros causa. Con esta información, el usuario podrá actuar de forma oportuna y tomar medidas frente a la problemática de siniestros viales.

En el mismo sentido, le permite identificar las carreteras con mayor tráfico y las relación que presentan con el número de siniestros, al mismo tiempo que le indica los factores más comunes contribuyentes de accidentes.

Por otro lado, un aporte muy importante del producto es que incluye un sistema de predicciones de accidentes viales basado en inteligencia artificial y aprendizaje de máquina, el cual le permite al usuario estimar para determinada fecha y Distrito de la Ciudad el número aproximado de accidentes. Esto es importante ya que permite tomar medidas correctivas antes de que sucedan los eventos mencionados, los cuales puede complementar con la información entregada por el Dashboard; a partir de ese análisis complementario puede llegar a concluir el lugar o carretera específica en el que se puede dar el accidente predicho y la causa más probable.

Ejemplo Ilustrativo:

A partir del análisis del Dashboard se puede observar que la vía Belt Parkway, es una carretera con alto número de siniestros viales y que de igual manera presenta tráfico elevado para el tercer trimestre del año 2022. Contrastándolo con el trimestre anterior, se refleja una disminución siniestral pero aún sigue siendo un número significativo. De acuerdo con la herramienta Dashboard se pueden identificar que los dos principales factores contribuyentes de los siniestros en esta vía son la distracción del conductor y conducción demasiado cerca con respecto al otro vehículo.

10.Manual de Instrucciones

Se cuenta con una página principal que muestra la portada, la cual le permitirá al usuario navegar sobre las diferentes visualizaciones que contienen información de interés, la primera de estas páginas entrega información sobre los Siniestros por Involucrado (Ciclistas, Conductores y Peatones).

En la siguiente página, se presenta la información en dos categorías: Siniestros leves refiriéndose a los heridos y graves refiriéndose a los muertos. La penúltima de las páginas dan información sobre Siniestros por Carretera (heridos y muertos), permitiéndole al usuario obtener datos de los siniestros sucedidos en las diferentes categorías mencionadas, y por último se encuentran los mapas con los que puede interactuar y ver la información de siniestros de forma resumida.

Portada:



Figura 4. Portada Dashboard

Interacción con la página:

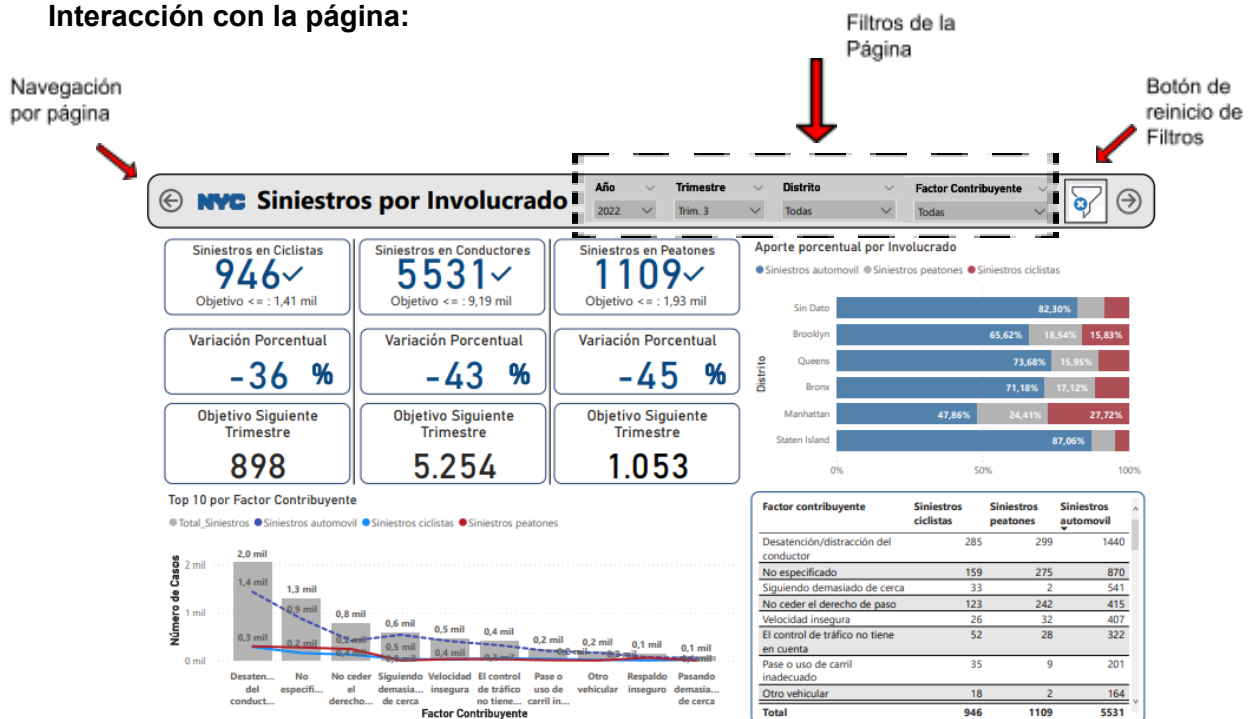


Figura 5. Interactividad y Filtros para Siniestros por Involucrados

Esta página cuenta con un panel de control con Botones de navegación para poder desplazarse de una página a otra, presenta un conjunto de filtros que le permitirá ubicarse en un periodo de tiempo año y trimestre específicamente. También le permitirá definir el distrito donde quiere ubicar los datos, y por último un filtro que le permitirá obtener los datos por factor contribuyente, es decir, la causa por la cual han ocurrido dichos siniestros. Las páginas cuentan con un botón que le permitirá reiniciar los filtros y volver al estado inicial.

Por otro lado, la página cuenta con un diagrama de barras 100% apiladas Superior derecha, con la cual se podrán ver datos del aporte porcentual de cada uno de los diferentes involucrados separados por Distritos de la ciudad de Nueva York.

En la parte inferior izquierda se encuentra un gráfico con el Top 10 de factor contribuyente y a la derecha de este una tabla que le permite ver todos los datos para el total de factores contribuyentes que se han registrado.



Figura 6. Representación e interpretación KPI'S

Cada KPI se encuentra dividida con por una línea recta que indica el sentido de la lectura de la información y están compuestas principalmente por:

- Una primera tarjeta que indica el nombre de la KPI, acompañada del valor obtenido para el trimestre actual, y el valor objetivo que se esperaba obtener. Siempre se busca que el objetivo para el siguiente trimestre sea menor o igual al indicado.
- La segunda tarjeta indica la variación porcentual KPI del valor actual con respecto al trimestre anterior, indicando en azul el cumplimiento del objetivo, y en rojo el no cumplimiento. Cada KPI, tiene como objetivo siempre lograr una disminución $\leq 5\%$ de siniestros según sea el caso.
- La tercera tarjeta indica el objetivo que se tiene que cumplir para el siguiente trimestre teniendo como referencia una disminución del 5% con respecto al valor actual. Opcionalmente, se encuentra otra tarjeta que indica el cumplimiento porcentual del periodo actual, teniendo como referencia un -5%.

A continuación se da un ejemplo de la correcta interpretación de una KPI:

- *Para la KPI 1, que corresponde con los siniestros (muertos y heridos) ocurridos a ciclistas, se puede observar que sí se cumplió el objetivo para el trimestre actual, ya que se esperaba tener menos de 1.41 mil siniestros y se obtuvieron 946 casos, con una disminución del 36% teniendo como referencia -5%. Por tanto, para el siguiente trimestre, el objetivo es lograr reducir los casos a 898 o menos, es decir, un 5% menos de los siniestros actuales.*

Interacción página Siniestros Leves y Graves

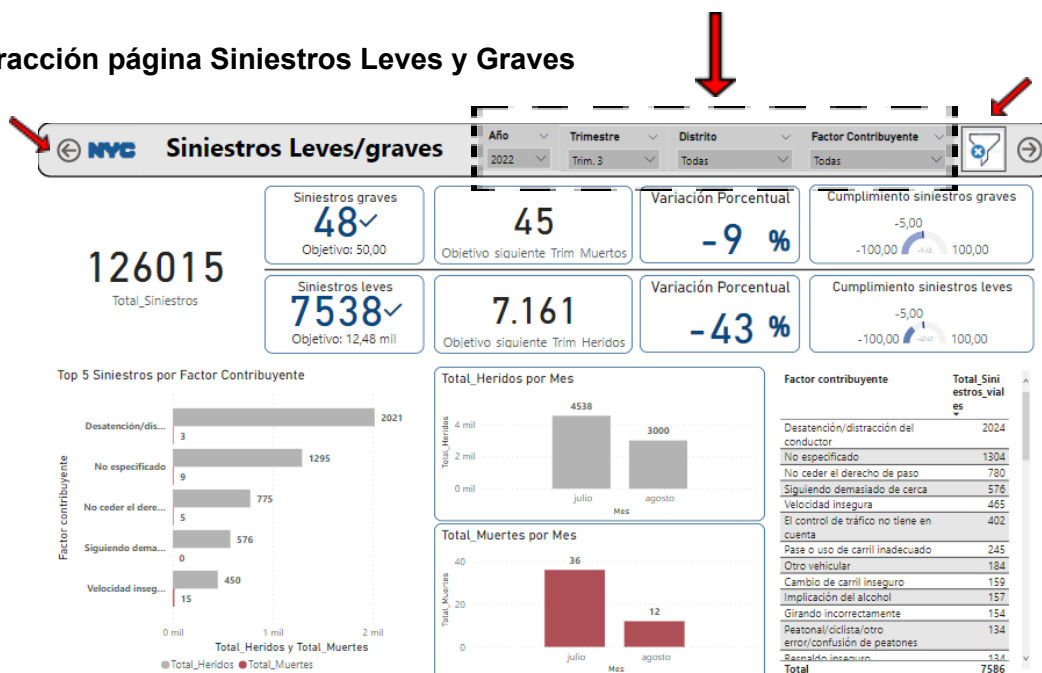


Figura 7. Interactividad y Filtros para Siniestros Leves y Graves

La interacción es similar a la anterior página, con sus respectivos botones de navegación, los filtros y el botón de reinicio de estos.

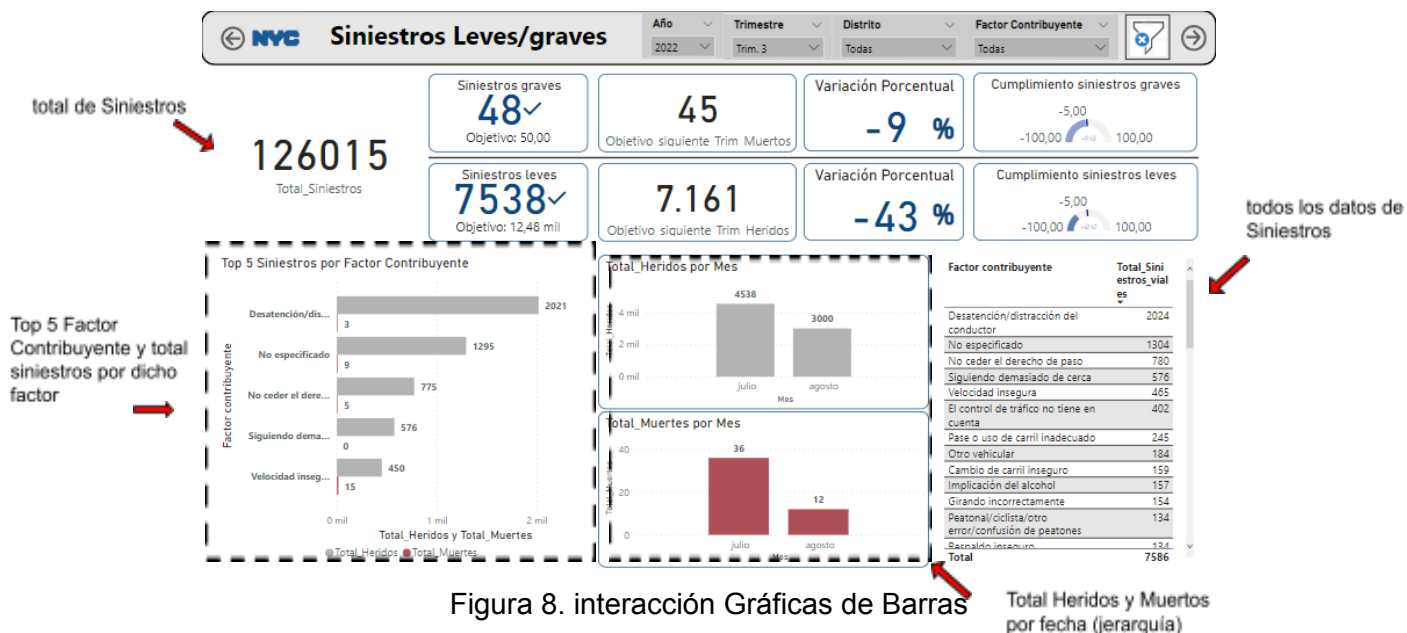


Figura 8. interacción Gráficas de Barras

Se presentan tres gráficos en la parte inferior de la página con lo cual se puede interactuar y ver los datos relacionados a los siniestros clasificados en dos categorías (graves y leves). El usuario puede observar el top 5 de siniestros relacionados con los primeros factores que contienen mayor número de estos casos. Por otro lado se encuentran en gráficas diferentes el total de siniestros por cada una de las categorías y se puede interactuar con los datos en

diferentes ejes de tiempo navegando por la jerarquía de fechas, y así obtener datos detallados en diferentes espacios de tiempo.

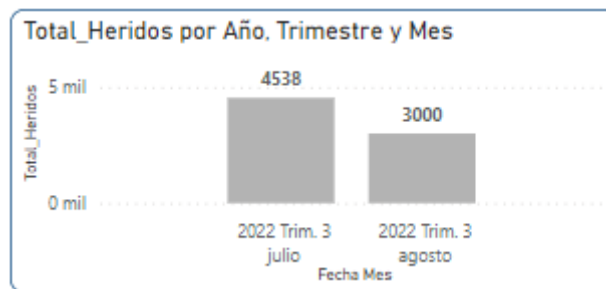


Figura 9. Interactividad por jerarquía del calendario

Le permite modificar el eje x de la gráfica en diferentes intervalos de tiempo, y en el ejemplo de la figura 9, en caso de haberse seleccionado un año en específico, la gráfica le permitirá bajar más en la jerarquía de fechas y ver los datos que se tienen para los meses relacionados a dicho trimestre y si así se desea a los días relacionados a estos.

Interacción página Siniestros por Carretera

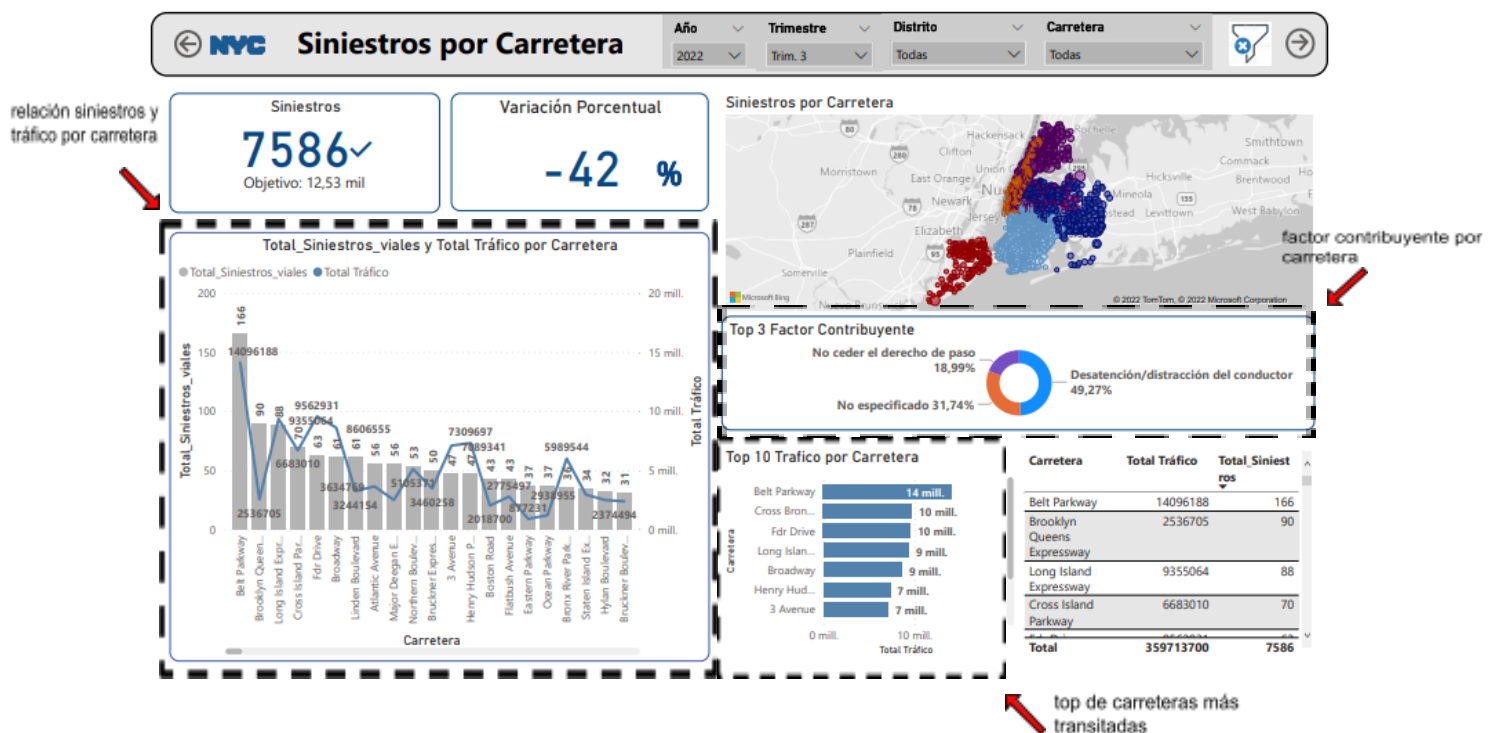


Figura 10. visualización Siniestros por Carretera

Su navegación es igual que las anteriores páginas, en este caso el objetivo de esta visualización es que el usuario pueda interactuar con los datos y ver las principales implicaciones de los siniestros en las diferentes carreteras dentro de la ciudad, igualmente le proporciona información de las principales causas de los siniestros relacionados a las carreteras que seleccione en específico (filtros) o de forma general. También le

proporciona información acerca del tráfico en dichas carreteras (total de vehículos o personas que pasan por el lugar) y que así pueda hacer un análisis con los datos que aquí se le proporcionan.

Interacción Página de Mapas

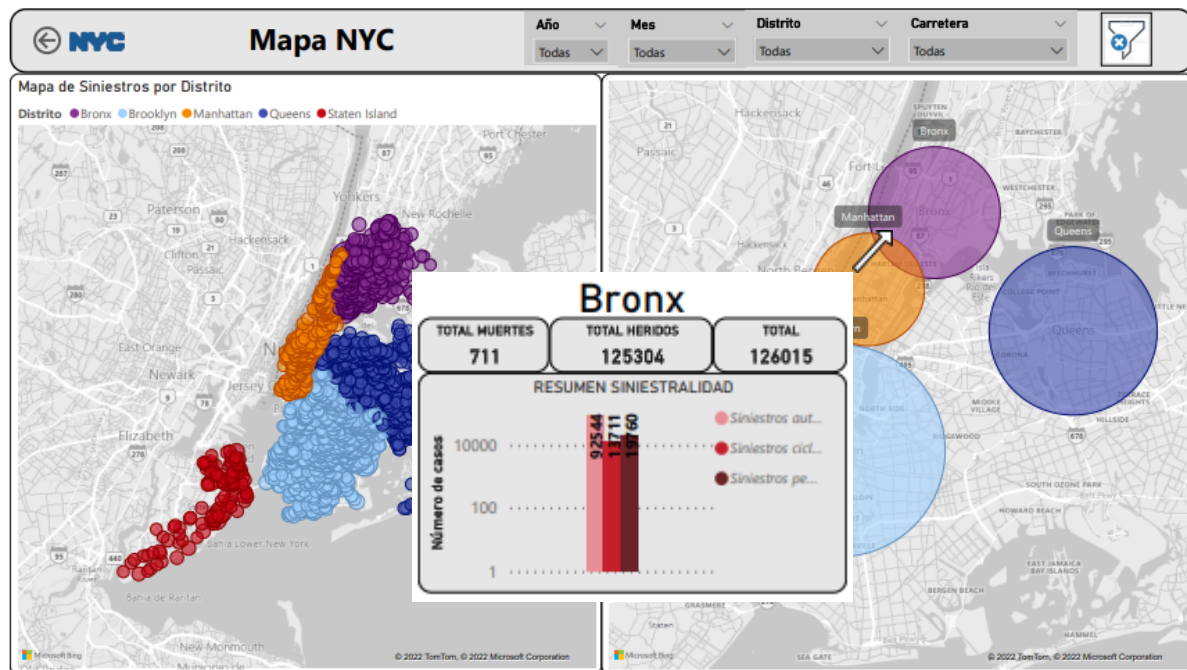


Figura 11. Interactividad Mapa de Siniestros

La última página corresponde al Mapa General de siniestros, en la cual se pueden encontrar los filtros de Año, Mes, Distrito y Carretera, adicionalmente un botón para reiniciar los filtros. Se encuentran dos tipos de mapa, en el derecho, se muestra una burbuja Grande de diferente color por cada Distrito, cuyo tamaño varía de acuerdo al número de siniestros (heridos y muertos) ocurridos durante el periodo seleccionado, por defecto, el acumulado de siniestros desde el 2020. Al pasar el puntero del mouse por una de las burbujas, se muestra una ventana de información que muestra un resumen con el número total de heridos y muertos, y un gráfico de barras por grupo (ciclista, peatón y automóviles). En el mapa de lado izquierdo, se puede observar de forma individual los distintos siniestros ocurridos de acuerdo a la ubicación en los diferentes Distritos, el cual permite identificar de forma precisa la dirección de dicho siniestro, además, cada Distrito, contiene múltiples burbujas identificadas según el color del distrito.

[illegible]

Referencias

[1]M. Ramírez, M. Salgado, H. Ramírez, E. Manrique, N. Osuna and R. Rosales, "Metodología SCRUM y desarrollo de Repositorio Digital", *Revista Ibérica de Sistemas y Tecnologías de Información*, 2018. [Accessed 19 August 2022].

[2]C. Rodríguez and R. Dorado, "¿Por qué implementar Scrum?", 2022.

[3] "El ABC de Trello: cómo utilizar tarjetas y tableros de Trello — Trello." [Online]. Available: <https://trello.com/es/guide/trello-1>