Propuesta de proyecto

1. **Entendimiento de la situación actual**

CASO

Se trata de una empresa de servicios de transporte de pasajeros que actualmente se encuentra operando en el sector de micros[[1]](#footnote-1) de media y larga distancia.

* La empresa está INTERESADA en invertir en el sector de transporte de pasajeros con automóviles[[2]](#footnote-2).
* La empresa QUIERE corroborar la relación entre dicho sector y la calidad del aire, como también la contaminación sonora, para ESTUDIAR la posibilidad de implementar vehículos eléctricos a su flota.
* La empresa PRETENDE hacer un análisis preliminar del movimiento de los taxis en NYC, para OBTENER un marco de referencia y poder TOMAR decisiones bien fundamentadas.

ROL A DESARROLLAR

El rol a desarrollar es el de un equipo de Data Science contratado para ACOMPAÑAR al negocio en dicho proceso de toma decisiones.

1. **Alcance**

El alcance de un proyecto como el que se plantea en el caso es muy amplio. Aun presuponiendo que la empresa ya ha desarrollado estudios de mercados para conocer la predisposición de los neoyorquinos a cambiar de proveedor de servicio de transporte de pasajeros, conoce las barreras de entrada y salida de este mercado, ha estudiado a la competencia, tiene un estudio de costos de operación, de inversión inicial, etc., el estudio para conocer la relación entre dicho sector y la calidad de el aire implicaría conocer y tener mediciones para las distintas fuentes de contaminación en la ciudad: calderas, otros motores, quema de basura, llantas, etc.; también implicaría relacionar una serie de contaminantes de los que no se tienen mediciones para toda la ciudad ni para cada hora o día. A su vez, para determinar con precisión habría que considerar la dirección del viento en el momento de la medición para cada estación de medición y el sentido del tráfico, entre otros factores.

En cuanto a decidir si se invierte o no en autos eléctricos por ‘una visión de un futuro menos contaminado’, implicaría como primer paso determinar la contaminación que se genera en la fabricación de un auto eléctrico y luego, una vez realizado un estudio a profundidad sobre el grado de contaminación de un vehículo de combustión, hacer una proyección a lo largo de su vida útil y compararlo contra lo que contaminó la fabricación del auto eléctrico y la generación de electricidad para cargar el auto.

En el presente estudio nos limitaremos a estudiar El PM 2.5 que es un tipo de contaminante atmosférico compuesto por partículas finas de diámetro menor o igual a 2.5 micrómetros. Estas partículas son emitidas principalmente por la quema de combustibles fósiles, procesos industriales y vehículos motorizados, y pueden tener efectos negativos en la salud humana al penetrar en los pulmones y causar problemas respiratorios y cardiovasculares.

Se decide por este contaminante porque es para el que se han encontrado más estaciones de medición en la Ciudad de Nueva York y; porque la medición del PM 2.5 tiene un alcance más reducido (de ‘neighborhood’ o barrio en español) que va de los 500 a los 4000 metros. Lo que permite encontrar una relación más directa con el tráfico de taxis en las zonas que abarca la medición.

Un estudio de CO, por ejemplo, con un alcance de (‘city’, ciudad) 4 km a 50 km, dada la ubicación de las estaciones de medición, estaría tomando (dejando de fuera la dirección del viento) fuentes contaminantes ubicadas fuera de la ciudad de Nueva York y dejando fuera fuentes contaminantes ubicadas en zonas de la ciudad como Jamaica Bay.

En cuanto a la contaminación sonora, que también le interesa a la empresa del caso, sería interesante el estudio ya que los autos eléctricos claramente tienen una ventaja en este rubro contra los de combustión. Abría que considerar estudios sobre accidentes ocasionados al no percibir el auto por la falta de sonido. Sin embargo, debido al margen de tiempo para desarrollar este proyecto y la dificultad para encontrar *datasets* adecuados con los que se pudiera cruzar información, se ha decidido dejarlo de lado.

Cómo segunda fuente, ha cruzar contra el *dataset* de taxis, se ha decidido por la medición de la precipitación (lluvia) en las diferentes zonas de Nueva York y estudiar su relación en el aumento o disminución en el número de viajes y en la duración de estos últimos.

El estudio está limitado a las zonas definidas por la “NYC Taxi & Limousine Commission”, a las que en adelante nos referiremos como zonas TLC. Estas zonas pueden estar compuestas por uno o barios barrios. En el repositorio de *github* puede consultarse en el notebook correspondiente a más detalle la composición de las zonas TLC.

1. **Objetivos**

Vamos a definir dos tipos de objetivos: a) los de nuestro equipo para este proyecto y; b) los que la empresa pudiera elaborar a partir de los nuestros.

1. Determinar la cantidad de viajes en taxi originados por Borough y por zona TLC.
2. Tener una participación del 10 % en todos los Boroughs y de al menos 5% en cada zona TLC.
   * 1. KPI Participación\_Boroughs = (cantidad de viajes por Borough de la empresa) / (cantidad de viajes por Borough totales)
     2. KPI Participación\_Zonas = (cantidad de viajes por zonas de la empresa) / (cantidad de viajes por zona totales)
3. Determinar el tiempo de espera promedio (para los VFH-HV) entre el momento en que se solicitó el servicio y el momento en que se recogió al pasajero; por Borough y por zona TLC.
4. Lograr un tiempo de espera promedio de 20 % o menor con respecto al promedio del sector en todos los Boroughs.
   * 1. KPI Tiempo\_espera = (Tiempo de espera promedio por Borough de la empresa) / (Tiempo de espera promedio por Borough del sector)
5. Determinar los Boroughs y zonas con mayores ingresos ponderados contra el número de viajes originados en dicho Boroughs y zona.
6. Tener una participación del 20 % en el Boroughs con más ingresos ponderados.
   * 1. KPI Mayores\_ingresos = (cantidad de viajes en Borough de la empresa) / (cantidad de viajes en Borough totales)
   1. Determinar el valor de la relación, si existe, entre la cantidad de viajes que inician y terminan, en el rango de una hora, en las zonas comprendidas en un radio de 2km a la redonda de la ubicación de cada estación de medición para el contaminante PM 2.5 medido por hora.
   2. Determinar el valor (medido en términos de aumento o disminución de solicitudes de servicios de transportes de pasajeros en vehículos) de la relación, si existe, entre que esté lloviendo y la cantidad de viajes en taxi.

1. **Documentación del proyecto**

Por el momento se puede consultar el avance y documentación del proyecto en el repositorio de github:

<https://github.com/oleaibarra/taxis_NYC>

Y en el drive de Google:

<https://drive.google.com/drive/folders/1-2C-7HrXhGXD_syOZdeJltS0-G0waiXm>

El stack tecnológico, la metodología, el diagrama de Gantt y otros documentos pueden ser consultados en el link proporcionado de Google drive.

La revisión de los *datasets* puede consultarse en el link the github proporcionado.

1. En algunos países, especialmente en América Latina, el término "micros" se utiliza para referirse a los autobuses de tamaño mediano que se utilizan para el transporte de pasajeros en distancias cortas y medias. Estos vehículos suelen tener capacidad para transportar entre 20 y 40 pasajeros y se utilizan comúnmente en áreas urbanas y periurbanas, así como en rutas regionales de media distancia. [↑](#footnote-ref-1)
2. El sector de "transporte de pasajeros con automóviles" se refiere a la industria del transporte que proporciona servicios de transporte de pasajeros utilizando automóviles, como taxis, vehículos de transporte con conductor (TNC) y servicios de limusina. Este sector incluye empresas y conductores que operan en una variedad de modelos de negocio, desde conductores independientes que poseen y operan sus propios vehículos hasta empresas que poseen y administran flotas de vehículos. [↑](#footnote-ref-2)