**MACHINE LEARNING SCIENTIST CAREER - DATACAMP**

**Enunciado Challenge N°3**

**Pronóstico de flujo vehicular en las estaciones de peaje de CABA**

**Challenge N°3:**

**Problema**

Pronóstico de flujo vehicular en las estaciones de peaje ubicadas en la Ciudad Autónoma de Buenos Aires.

**Asignación**

Diseñar y desarrollar un modelo de forecasting basado en técnicas estadísticas, modelos de series de tiempo o de machine learning para pronósticar el flujo vehicular en 6 estaciones de peaje ubicadas en la Ciudad Autónoma de Buenos Aires.

**Ventana de evaluación**

El horizonte de forecasting a utilizar como “datos desconocidos del futuro” está definido por los últimos 6 meses anteriores al inicio del aislamiento obligatorio que tuvo lugar en Marzo de 2020. Como se supone que estos valores son completamente desconocidos, dicha ventana de tiempo NO puede ser utilizada para el entrenamiento, selección o testeo de modelos.

**Fair play**

Para garantizar la transparencia de la competencia, una persona ajena a los participantes elegirá una ventana de tiempo con inicio y fin desconocidos para estos, pero contenida dentro de los 6 meses de la ventana de evaluación descrita anteriormente. De manera que no sea posible conocer el intervalo exacto, y por ende overfittearlo, sobre el cual se definirá el ganador de la competencia.

**Condiciones**

* **Lenguaje:** Python.
* **Tiempo:** 2 semanas (acordar inicio).
* **Métrica:** WAPE a nivel general de los forecasts realizados para la ventana de fair-play contenida dentro de la ventana de evaluación.
* **Estaciones a considerar:** Alberdi, Avellaneda, Illia, Retiro, Salguero y Sarmiento.

**Requerimientos**

Cada participante debe incluir en su solución del problema de forecasting como mínimo los siguientes elementos:

* Identificación y corrección de outliers (si existen)
* Imputación de valores ausentes (si existen).
* Plantear un enfoque para la validación y selección de modelos acorde a la naturaleza del forecasting.
* Una muy breve PPT donde se muestre la metodología seguida para la solución del problema.

**Evaluación y premiación**

El ganador será elegido como el modelo con la mejor performance (WAPE) a nivel general para la ventana de fair-play dentro del intervalo de evaluación.

El premio que recibirá el ganador será la posibilidad de elegir un lugar para ir a cenar y ser invitado por los otros participantes a dicho lugar sin tener que aportar para el pago de la cuenta.

**Datasets**

Los datasets a utilizar son los correspondientes a los años 2016, 2017, 2018, 2019 y 2020 de los datos de flujo vehicular publicados por el gobierno de la ciudad, los cuales se pueden encontrar en el siguiente link: https://data.buenosaires.gob.ar/dataset/flujo-vehicular-por-unidades-peaje-ausa