# Neural Network Assignment

## Scenario

Launched in Spring 2015, Indego - Philadelphia's bikeshare program – offers self-service bikes via over 100 stations all day, every day. Philadelphia's city government owns bicycles and stations, with the Mayor's Office of Transportation and Utilities (MOTU) planning and managing the project. Indego rests on a public-private partnership that includes Bicycle Transit Systems, which provides maintenance, marketing, and customer service as well as sponsorship by Independence Blue Cross. This alternate form of public transportation increases the mobility of Philadelphians to get where they want to go, on their schedule, while providing more opportunities for healthy physical activity. Individuals can buy a monthly or yearly pass or purchase a trip at any station with a credit card. Visit **RideIndego.com** for more information on how the program works.


You have been asked by the City of Philadelphia and its partners at Bicycle Transit Systems to build a model that will forecast Indego ridership based on historical data. Ridership is measured by the number of trips taken on a given day.

## Data

**https://raw.githubusercontent.com/SueMcMetzger/MachineLearning/main/chpt10/IndegoRides.csv**

Indego trip data that has been downloaded from OpenDataPhilly.org and summarized for the purpose of this assignment. The aggregated daily data is very clean and includes for each day the average number of trips taken, the average duration of bike usage on a given day, the type of bicycle (Standard vs Electric), the type of trip (One Way vs Round Trip), as well as weather details for the day. The weather data includes the average humidity, the average maximum and minimum temperatures, as well as a binary value for report severe weather conditions (like snow, wind or rain).

As can be imagined, ridership patterns will see changes by the day of the week as well as the seasons. For that reason, it is recommended you add into your model the month of the year (as categorical) as well as the year (ordinal). It will be helpful to add a feature that identifies the day of the week (as categorical). Feel free to consider additional data to support this model.

## Instructions

Using Python with Keras, complete the following:

1. Build a neural network that can forecast the number of trips.
    a. Since average duration is correlated with the number of trips, the average duration should be removed from the model.
    b. Determine a naïve baseline prediction as a litmus for model performance.
    c. Tune this neural network using both your intuition and the hyperband.

d. Using your best model, predict the number of trips for the 5 dates provided in the file IndegoRidesToPredict.csv.

2. Build a neural network that can forecast the anticipated average duration for any given day.
    a. Duration should be defined as one of 3 types of days:
        i. Quick (an average duration 25 minutes or less)
        ii. Average (an average duration of more than 25 minutes but less than 55 minutes)
        iii. Cruising (an average duration of more than 55 minutes)
    b. Since the average duration is correlated with the number of trips, the number of rides should not be included in the model.
    c. Determine a naïve baseline prediction as a litmus for model performance.
    d. Tune this neural network using intuition and the hyperband.
    e. Using your best model, predict the anticipated duration for the 5 dates provided in the file IndegoRidesToPredict.csv.

## Submission

Complete the presentation file that highlights what is important when sharing a model with others. This includes data preparations, model design, model performance and predictions. Since this is a professional document, see that the structure and your analysis is concise and direct (bullet points rather than paragraphs). Save this presentation as PDF to ensure it preserves all formatting and visuals accordingly. Submit to Blackboard your **presentation**, a link to your Google Colab **notebook file** (.ipynb), as well as a **PDF of your notebook**. See that the Jupyter notebook files include the most current output for each cell (this aids the instructor).

You will be evaluated on your ability to analyze data, implement & tune machine learning models using Python, and assessment of the model results.