**Capstone Project Proposal:**

**TimelyTravels: Flight Delay Prediction using Machine Learning**

Flight delay can be a cause of frustration amongst the travellers. I want to build a model that can predict the flight status of a given flight based on certain features.

1. Data Gathering:

* I have gathered 1 year data from BTS-United States Department of Transportation (<https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FGK&QO_fu146_anzr=b0-gvzr>)
* The data contains US domestic commercial flights from May 2022-Apr 2023.
* There are more than 7 million rows and 120 columns.

1. Data Cleaning and Pre-Processing:

* Read the monthly csv files from different paths and combine them.
* Drop duplicate, irrelevant, redundant and high missing value columns.

1. Data Visualisation

* Visualise the statistics and trends in the data.
* Create graphs and charts like:
* Is there a monthly or weekly dip in flights?
* Which Airline is performing how?
* How many flights are on-time?
* What is the common/major cause of delay?

1. Building Classification Models

* Since the data has columns with delay details, I can classify the data in On-time(0) and Delayed(1) flight statuses.
* I plan on building 3-5 classification models

1. Selecting Best based on Model Validation

* Compare the confusion matrix, accuracy scores and classification reports of the models built and select the best model for deployment.

1. Deploying the project

* Deploy using Flask app.
* The model takes in values of flight details and gives the output delayed or on-time.