**Capstone Project 2: Project Proposal**

**Title: New York City taxi fare prediction**

1. **Business Problem**:

In this problem, we will predict the fare amount (inclusive of tolls) for a taxi ride in New York City, given the pickup and dropoff locations. The price estimation based on the distance between the pickup and dropoff locations results in an RMSE of $5-$8. The proposed Machine Learning model should obtain better results than the price estimation based on the distance between two points.

1. **Business Objective**:

Build a machine learning model that gives an error rate at least less than 5.0 RMSE to predict New York City taxi fare based on two given points.

1. **Data**:

The data contains three files, and the training data file has eight features that provide the details of yellow taxi rides in New York City.

Data Source: <https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/data>

Features:

* key: ID variable that provides a unique string to identify each row.
* fare\_amount: Target variable that provides the amount of the taxi ride in dollars.
* pickup\_datetime: It provides the start of the taxi ride as a timestamp.
* pickup\_longitude: Longitude coordinate of the pickup location.
* Pickup\_latitude: Latitude coordinate of the pickup location.
* dropoff\_longitude: Longitude coordinate of the dropoff location.
* dropoff\_latitude: Latitude coordinate of the dropoff location.
* passenger\_count: The number of passengers in the taxi ride.

1. **Methodology**:

The New York City Taxi fare prediction is a supervised regression problem, and we will explore the data to choose the right regression model. In this project, we will follow these steps:

* Problem statement: Business problem and objective
* Data wrangling: Data cleaning and Exploratory Data Analysis (EDA)
* Feature engineering
* Model evaluation
* Model validation and analysis

1. **Deliverables**:

The deliverables will contain the complete project code along with the final report and presentation slides stored in a Github repository.