**Question/need:**

* What is the framing question of your analysis, or the purpose of the model/system you plan to build?

The purpose of this model is to help taxi drivers and the city of NYC better understand why passengers tip/what factors tips are influenced by.

* Who benefits from exploring this question or building this model/system?

NYC taxi drivers.

**Data Description:**

* What dataset(s) do you plan to use, and how will you obtain the data?

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Taxi trip records are available here by month as csv files.

* What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

The features in this dataset include: pickup datetime, dropoff datetime, passenger count, trip distance, payment type, fare amount, tax, tip amount, extra charge, toll amount, total payment, and congestion surcharge. An individual row of data will be one unit of analysis.

* If modeling, what will you predict as your target?

I will be predicting if the passenger tips or not. (If tip = 0 or doesn’t.)

**Tools:**

* How do you intend to meet the tools requirement of the project?

I will use a Python package to create and interpret my classification model.

* Are you planning in advance to need or use additional tools beyond those required?

If possible, I’d like to create some visualizations in d3 or Tableau! I’ve also discussed creating a more interactive app with my data

**MVP Goal:**

* What would a [minimum viable product (MVP)](https://github.com/thisismetis/Metis_Fundamentals/blob/main/project_deliverable_templates/mvp.md) look like for this project?

An MVP could be initial visualizations on the relationship between the features on the target, or chosen classification metric values from the baseline/not-fully-refined model.