37. Caesar’s Taxi Prediction Service

- **Report Link**: PDF
- **Year**: Fall 2018 (CS229)

1. Problem Statement & Formalization

Predict taxi trip demand / ride requests (or service) in a city.

- **Input**: spatio-temporal features (time, location, historical demand)
- **Output**: predicted number of taxi requests or recommended dispatch actions
- **Goal**: anticipate taxi demand to position taxis efficiently
- **Metrics**: RMSE, MAE, possibly accuracy of high-demand area predictions

2. Data / Dataset

- Taxi trip logs (historical pickup counts, time & location) — used by the report.
- Public equivalent: New York City Taxi & Limousine Commission (NYC TLC) dataset

3. Model Design / Methods

- Baseline: linear regression or Poisson regression
- Advanced: decision trees, gradient boosting, time series models (ARIMA, LSTM)
- Feature engineering: time-of-day, day-of-week, spatial bins, lagged demand

4. Training & Validation

- Split data by time (train on past, test on future)
- Cross-validation on temporal folds

- Hyperparameter tuning

5. Evaluation & Reporting

- Report RMSE / MAE of demand predictions
- Compare baseline vs advanced models
- Show spatial maps of prediction vs truth

6. Extensions / Bonus

- Real-time prediction (streaming)
- Incorporate external features (weather, events)
- Prescriptive dispatch: suggest taxi movement/placement

7. Deliverables

- Taxi demand dataset subset
- Prediction code
- Report with metrics + maps of predictions