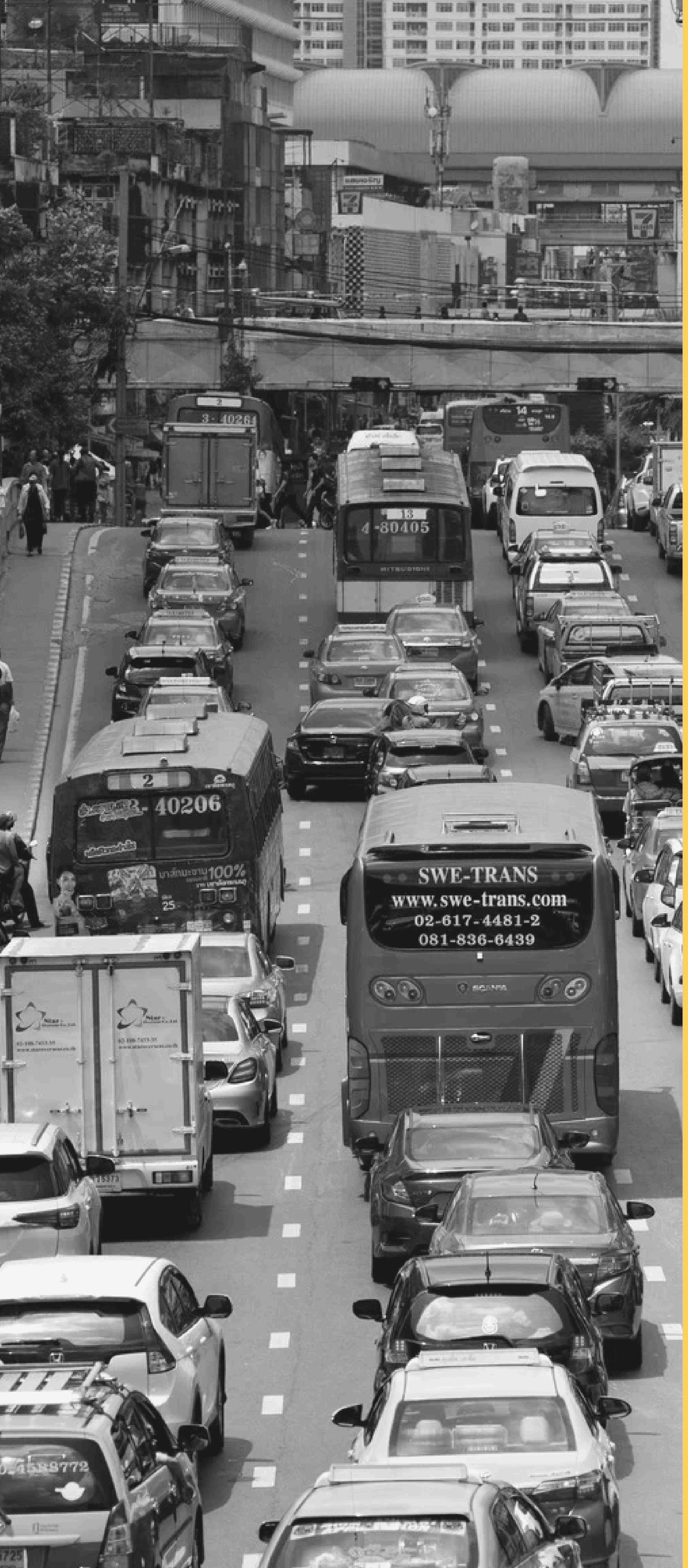
TAXI DEMAND FORECASTING

Ngoc Nguyen, Edward Ramirez, Sriram Raghunath,
Nazanin Komeilizadeh, Noah Gillespie, Li Meng

GitHub Repo



Motivation

- Reliably predicting taxi demand in real-time can vastly empower on-demand movements across the city:
  - Taxi companies can dispatch drivers promptly
  - Drivers can optimize route decisions to maximize earnings
  - Customers are more likely to experience timely service
- Our project: predict taxi demand with **deep learning models for times series**
  - Multilayer Perceptrons (MLP)
  - Long Short-Term Memory (LSTM)
  - Temporal Graph-based NN

Data and Preprocessing

- Raw Data: NYC TLC Monthly Trip Record Data
  - Records of individual routes
    - Pickup/Dropoff Location (Taxi Zones)
    - Pickup/Dropoff Times
    - Pricing information, trip distance, etc.
  - Yellow taxis, ridesharing vehicles, and others
  - Since 2009

- Processed Data*:
  - Hourly Time Series Data for each Pickup Location
    - Number of rides per hour
    - Average price of rides per hour
    - Manhattan Taxi Zones
  - Yellow Taxis
  - 01/01/2022 - 03/31/2024

	tpep_pickup_datetime	tpep_dropoff_datetime	PULocationID	DOLocationID	total_amount	tip_amount
0	2023-01-01 00:32:10	2023-01-01 00:40:36	161	141	14.30	0.00
1	2023-01-01 00:55:08	2023-01-01 01:01:27	43	237	16.90	4.00
2	2023-01-01 00:25:04	2023-01-01 00:37:49	48	238	34.90	15.00
3	2023-01-01 00:03:48	2023-01-01 00:13:25	138	7	20.85	0.00
4	2023-01-01 00:10:29	2023-01-01 00:21:19	107	79	19.68	3.28

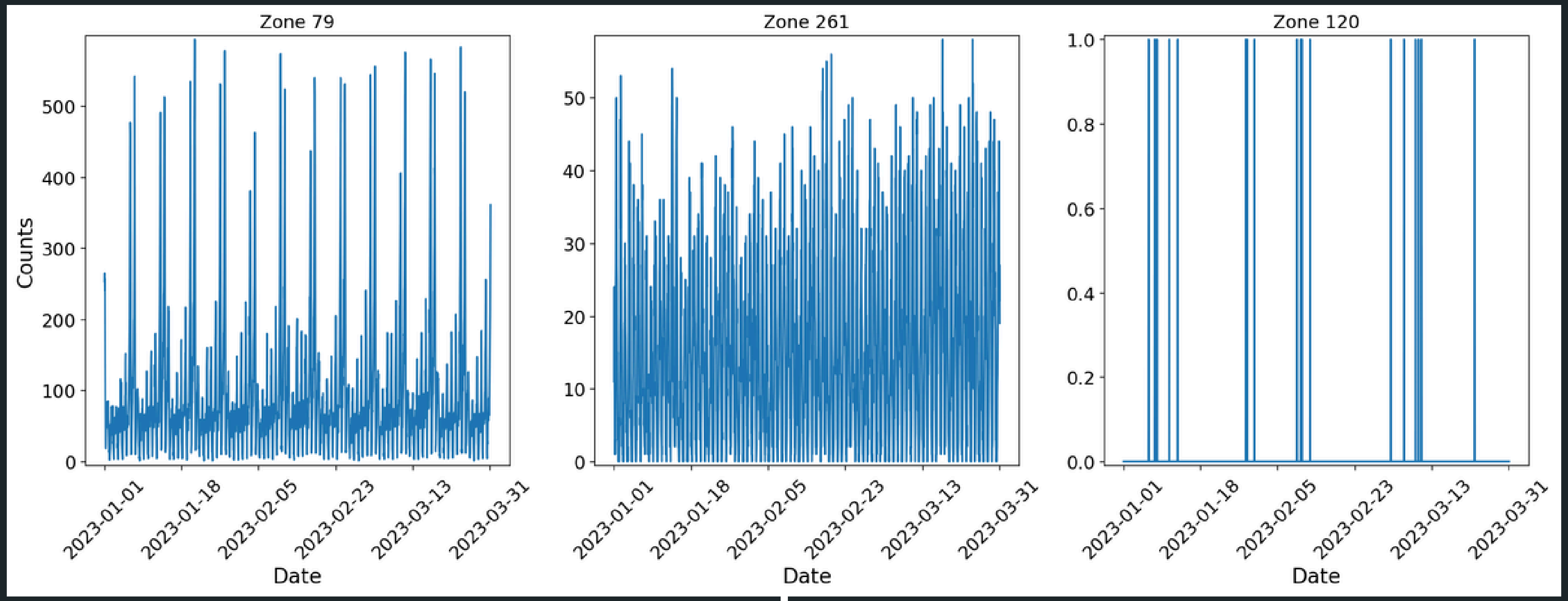
NYC Taxi & Limousine Commission

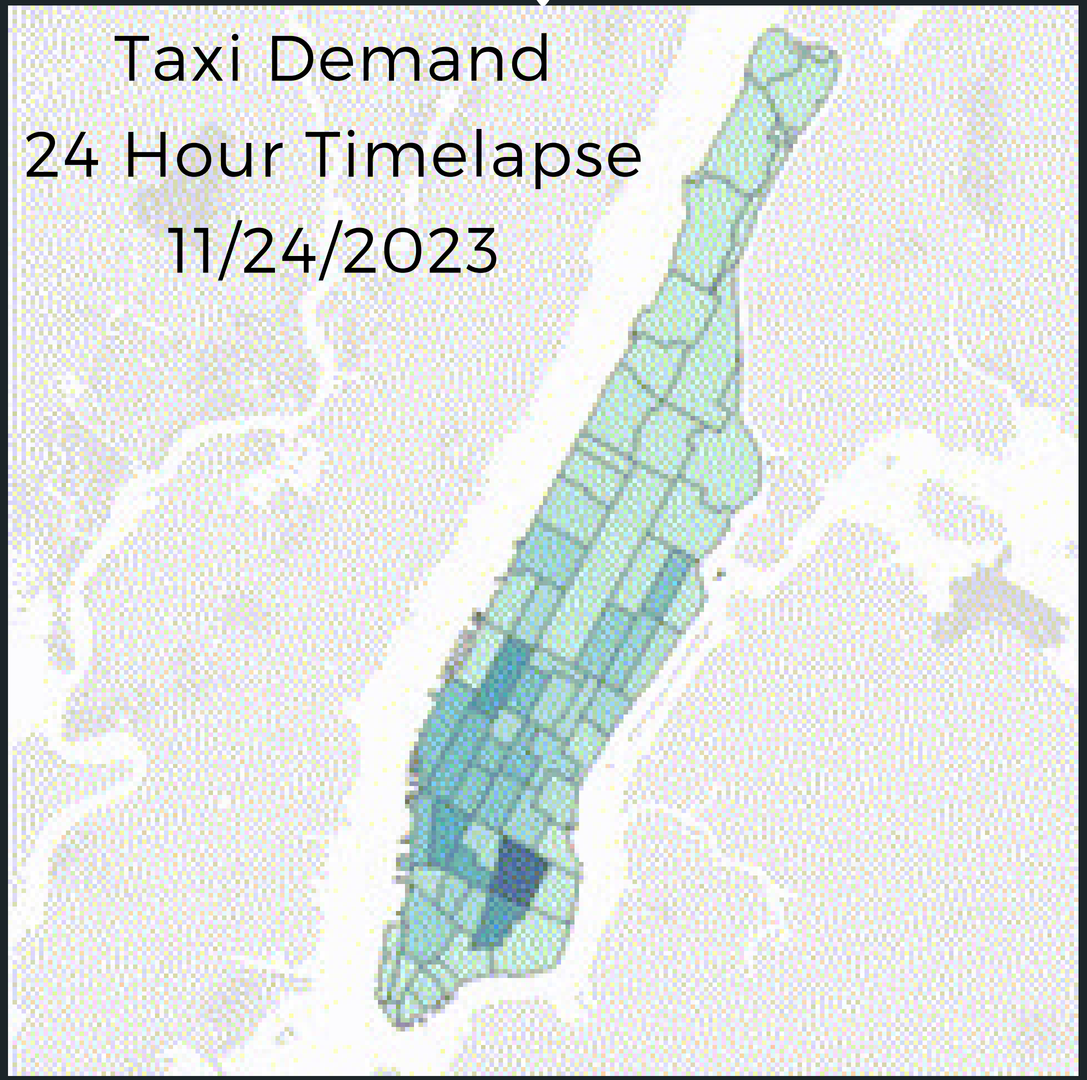
*For further details on processing, see the corresponding GitHub repository

Data and Preprocessing

- Raw Data: NYC TLC Monthly Trip Record Data
  - Records of individual routes
    - Pickup/Dropoff Location (Taxi Zones)
    - Pickup/Dropoff Times
    - Pricing information, trip distance, etc.
  - Yellow taxis, ridesharing vehicles, and others
  - Since 2009

- Processed Data*:
  - Hourly Time Series Data for each Pickup Location
    - Number of rides per hour
    - Average price of rides per hour
    - Manhattan Taxi Zones
  - Yellow Taxis
  - 01/01/2022 - 03/31/2024

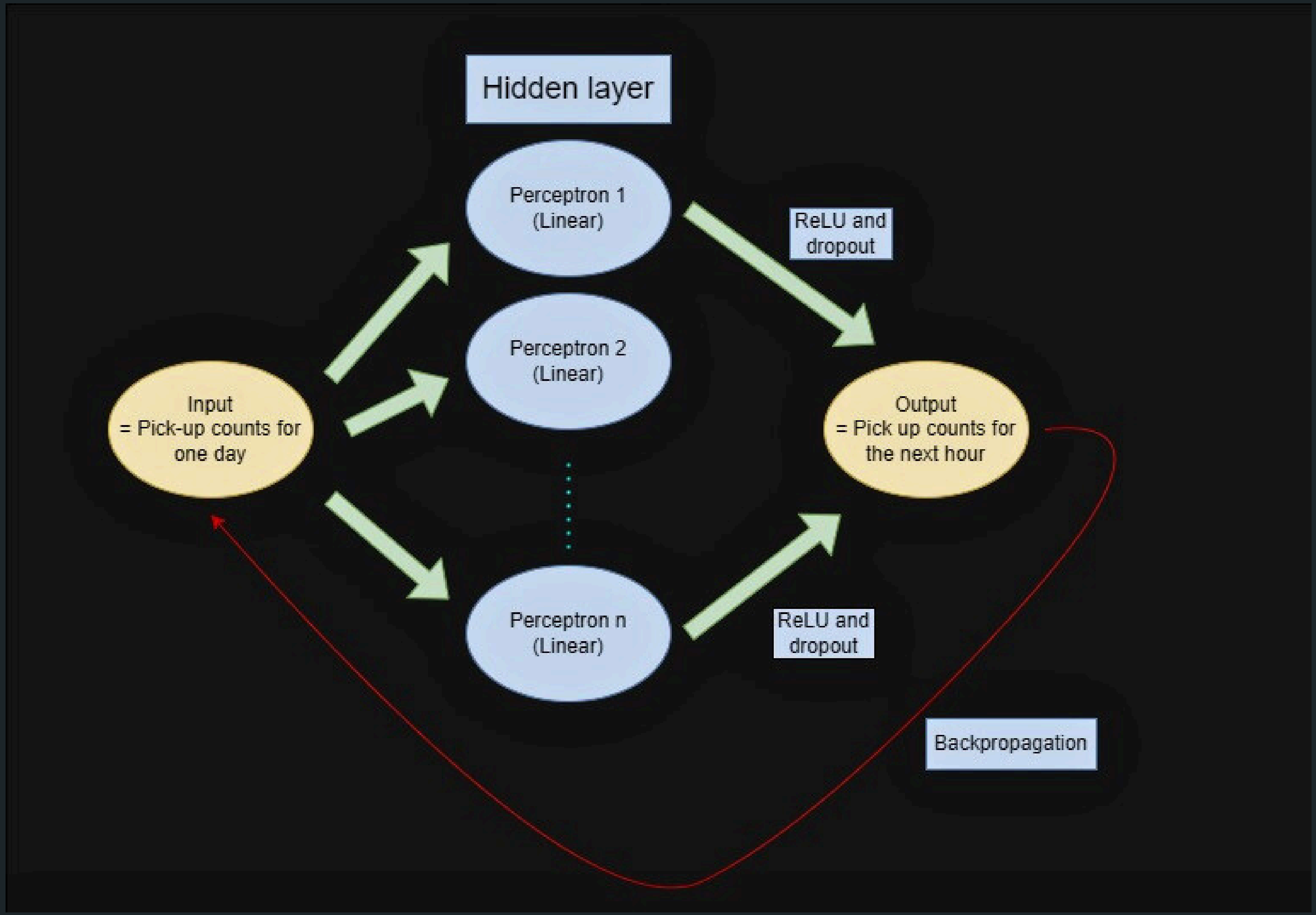




*For further details on processing, see the corresponding [GitHub repository]

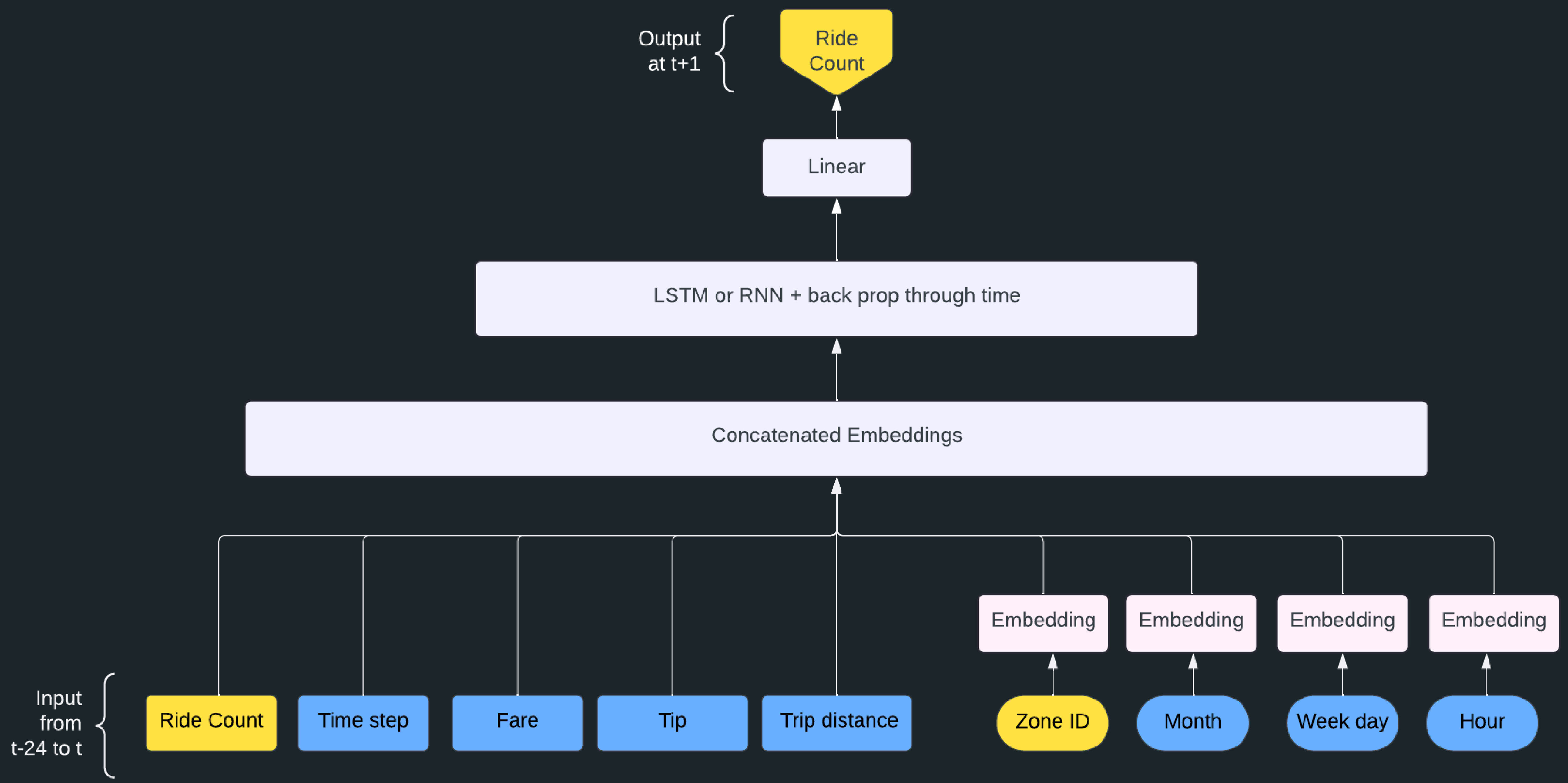
Modelling: Multilayer Perceptrons

- Simple feed forward neural network with one hidden layer
- Sliding window technique uses window size of one day to forecast next hour
- Hyperparameter tuning to obtain best learning rate, scheduling, number of neurons in hidden layer, dropout probability



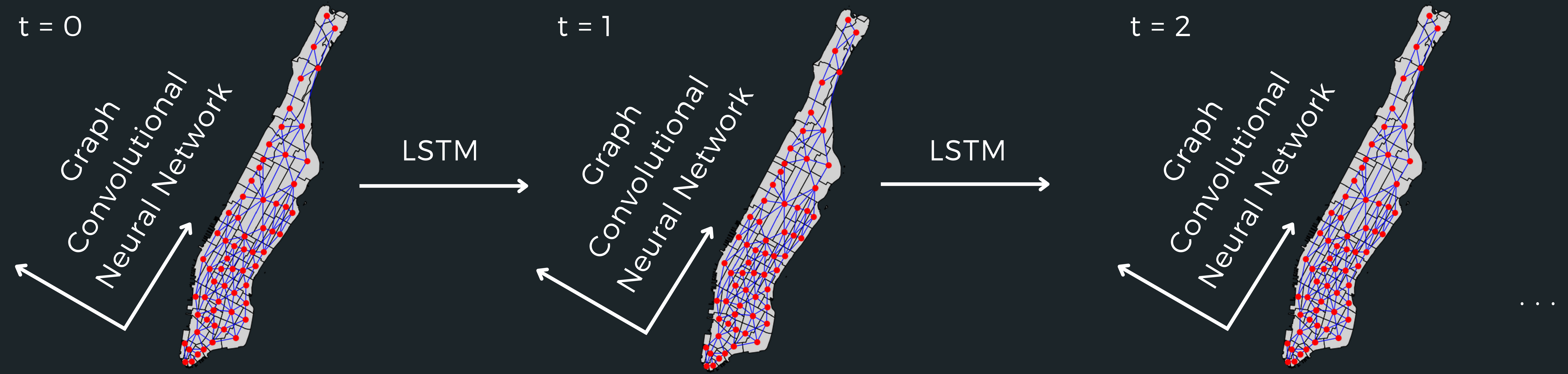
Modelling: Multi-series LSTM Models

- Fit taxi counts time series for all zones simultaneously with backpropagation, moving through time in 24-hour steps.
- Categorical embeddings: help model learn each zone's unique characteristics
- Multivariate version: add additional features to capture both historical trends and seasonality patterns



Modelling: Temporal Graph-based NN

- Add graph structure to LSTM: **temporal** pattern within zone + **spatial** correlation between zones
  - Graph nodes: 63 taxi zones in Manhattan
  - Edges of the graph: distances between the taxi zones
  - Use taxi counts as the node features
- Predict demand for the next hour based on the previous 24 hours of data for each zone

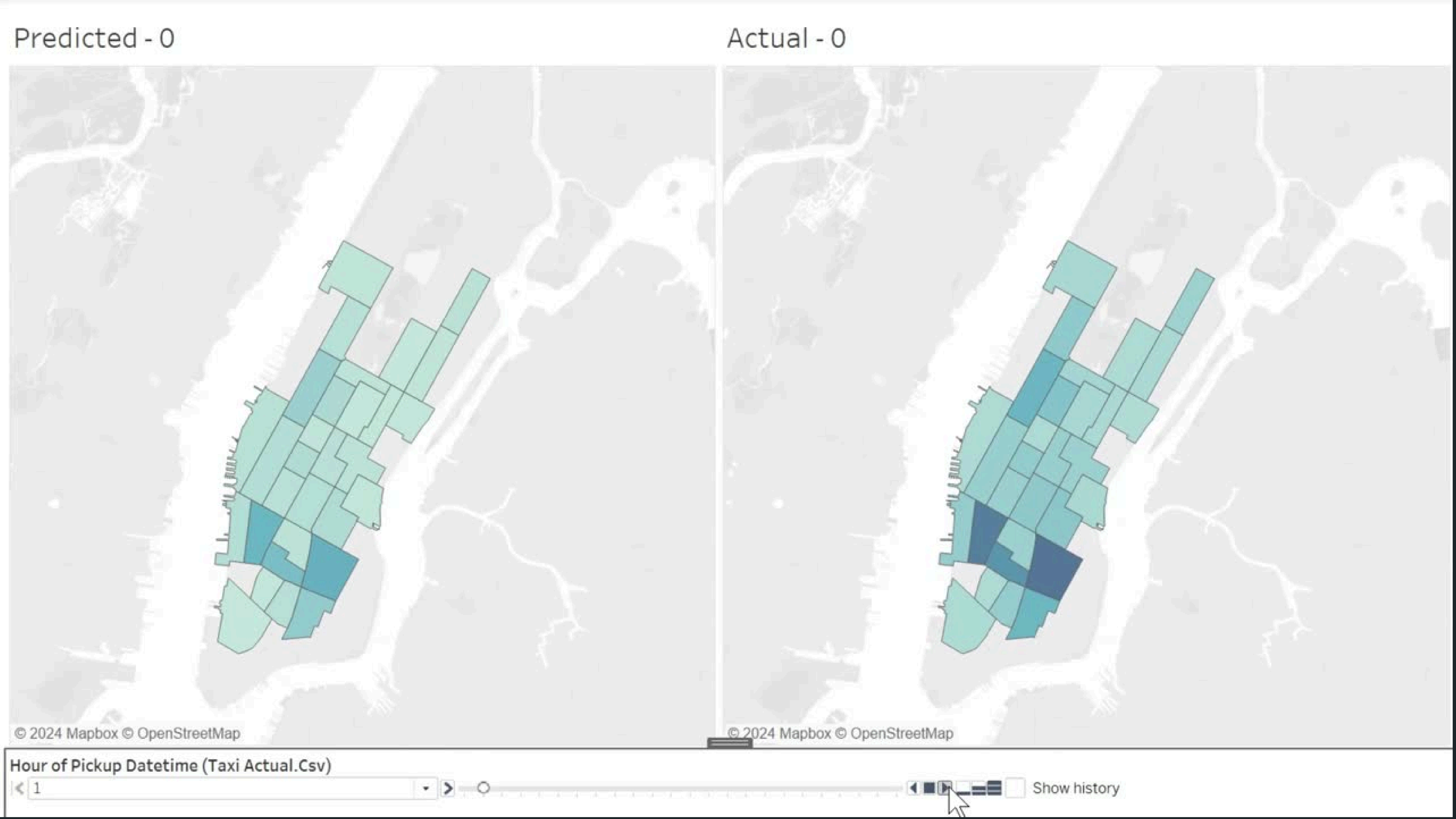


Model Evaluation

- Baseline: ARIMAX
  - ARIMA models the random nature of the time series.
  - A Fourier Series models the daily and weekly trends of the time series.
- Multilayer perceptron performs better than ARIMAX
- Temporal models performed the best
  - Including fare information made the predictions worse
  - Time variables that encoded day of week, month of year, and hour improved performance

Model	**KPI (Validation Set)**	
	RMSE (Rides)	SMAPE
ARIMAX	40.02	69.57%
Multilayer Perceptrons	23.58	63.37%
Ride Counts LSTM	16.56	57.82%
Ride Counts + Time Vars LSTM	15.86	57.81%
Multivariate + Time Vars LSTM	16.46	57.25%
Temporal Graph-based NN	17.09	55.38%

Graph Neural Network Predictions



Future Directions

A few directions that can enhance our best model and extend its application:

- Model to predict demand of particular routes
- Predict the next few hours, as opposed to the next hour
- Additional features outside the dataset such as weather forecasts
- Larger training dataset: more taxi zones + ridesharing vehicle data
- Hyperparameter tuning