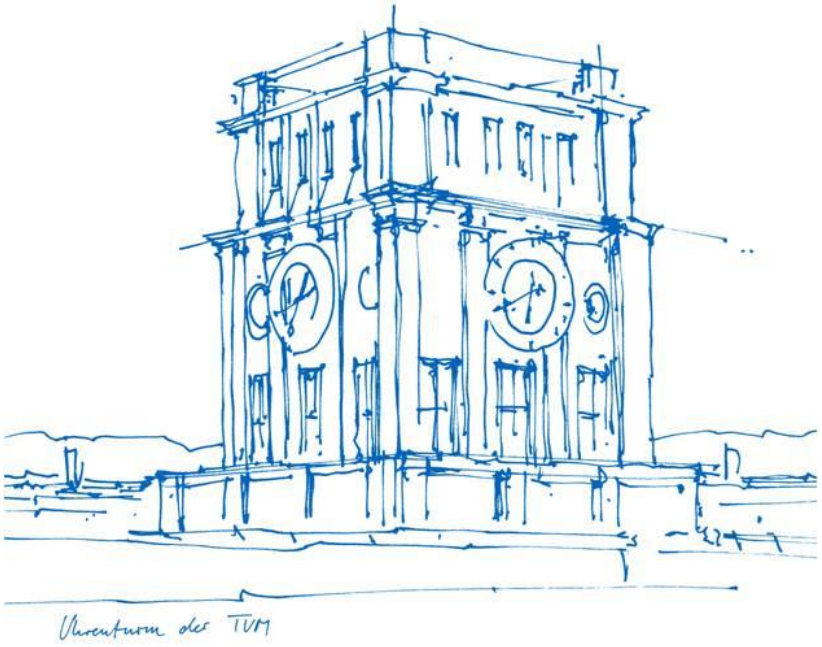
DMLab Winter 17/18
Green Taxi Trip Dataset

Taxi Mobility Pattern

Ruidong Zhang

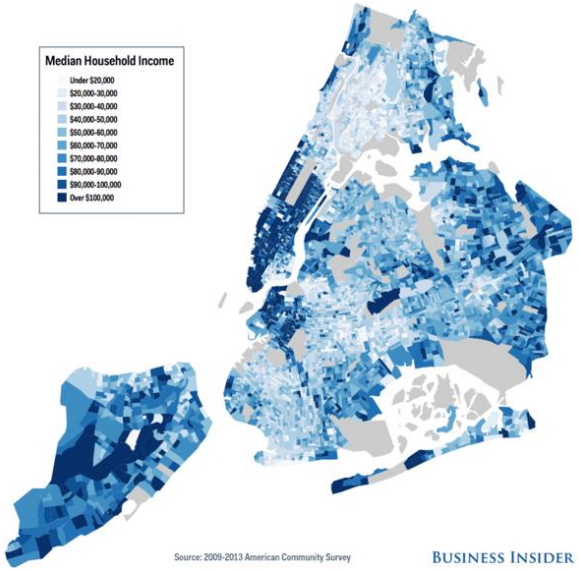


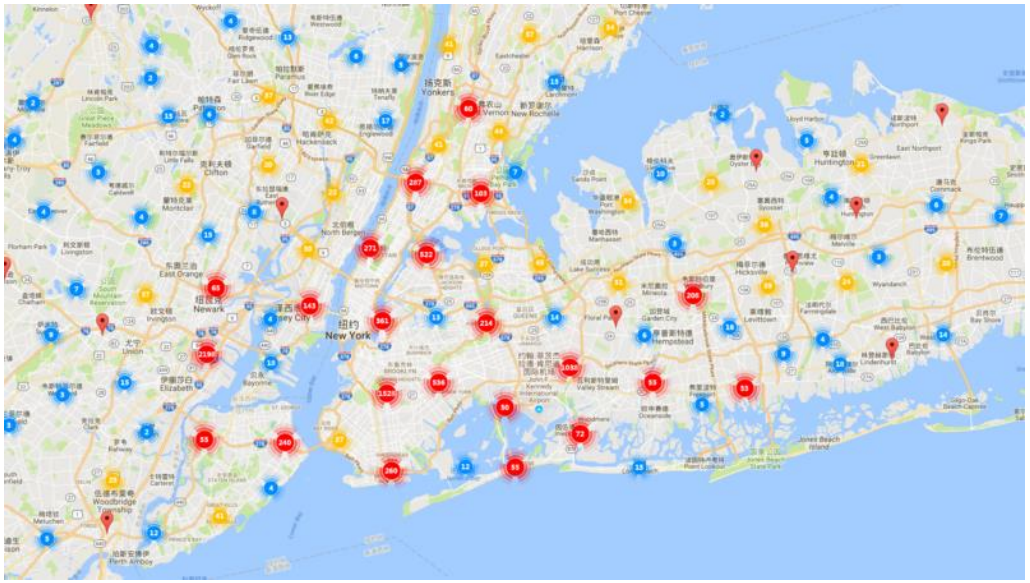
Dataset

- Size: ~ 2.1 GB, 16.4 million rows
- 23 features, dropped 3

Relation between Location and Amount charged

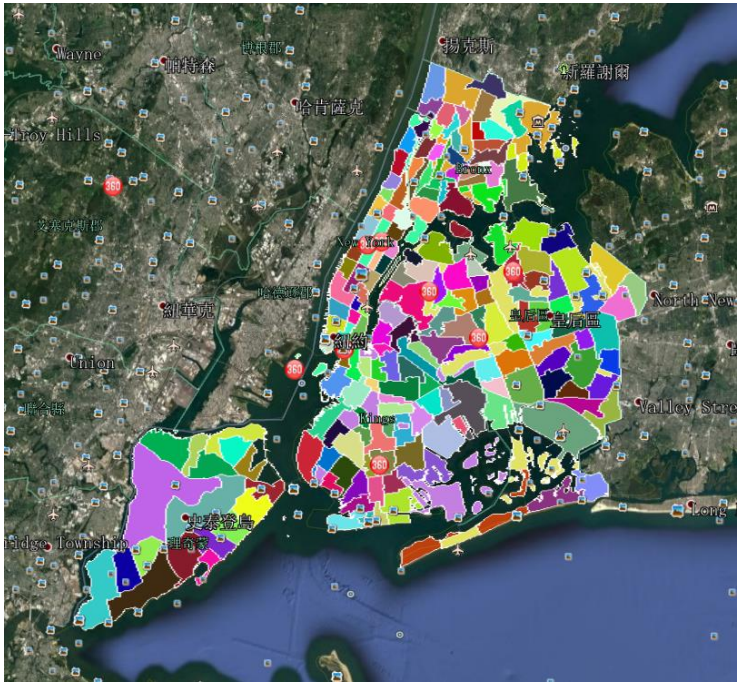
- ➔ Cluster & plot in Google Map
- ➔ Filter records by total amount
- ➔ Hotspots for high fare amount = areas with higher income
- ➔ Wise to wait at dark area





Mobility Pattern Discovery & Forecasting

- Map location into districts
- Create transition matrix with framework

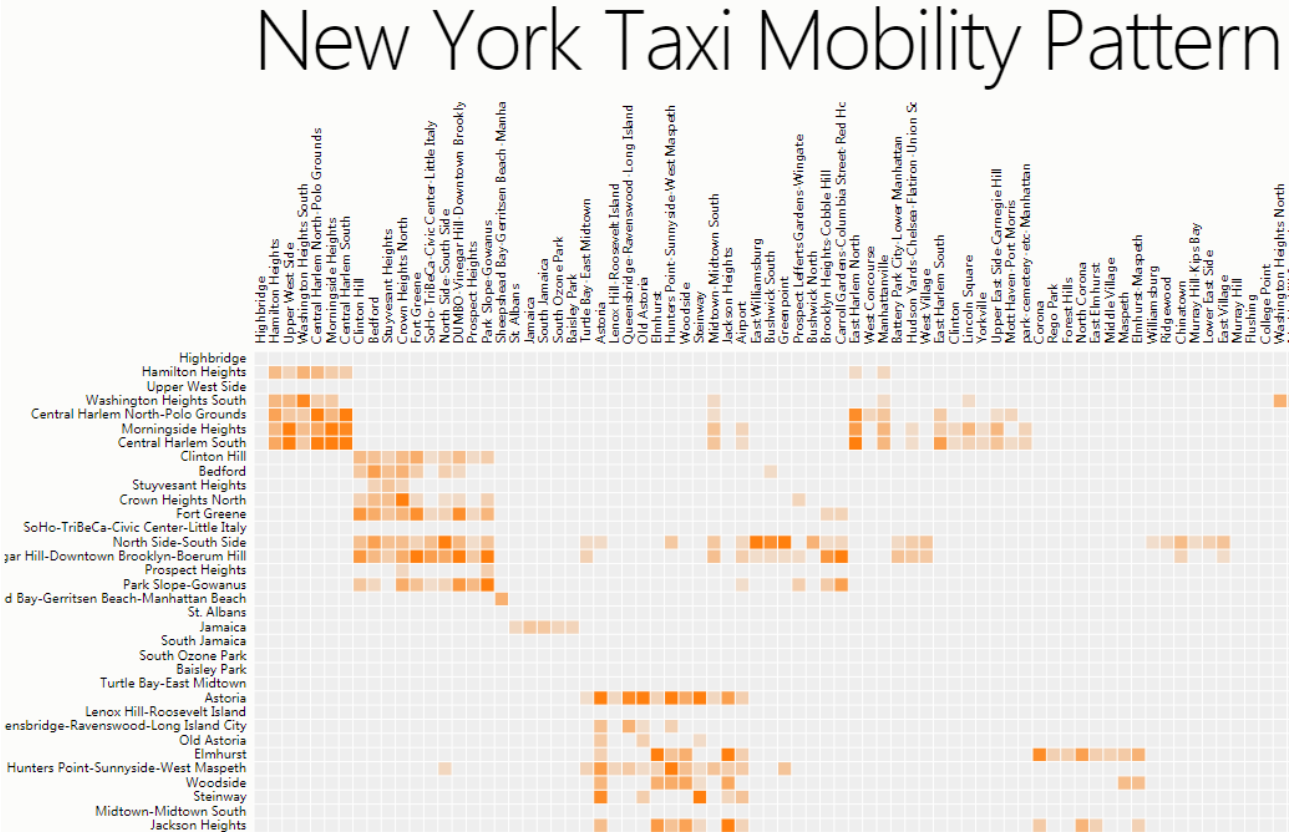


GET /rest/matrixedges/?tripset=29&weight__gte=5000&limit=10&ordering=-weight


```
{
    "source": "Astoria",
    "target": "Astoria",
    "value": 110828.0
},
{
    "source": "North Side-South Side",
    "target": "North Side-South Side",
    "value": 103014.0
},
{
    "source": "Park Slope-Gowanus",
    "target": "Park Slope-Gowanus",
    "value": 92105.0
},
{
    "source": "North Side-South Side",
    "target": "Greenpoint",
    "value": 77013.0
},
{
    "source": "Central Harlem South",
    "target": "Central Harlem North",
    "value": 76009.0
},
{
    "source": "East Harlem South",
    "target": "East Harlem North",
    "value": 71337.0
},
{
    "source": "Jackson Heights",
    "target": "Jackson Heights",
    "value": 67518.0
},
{
    "source": "Central Harlem North",
    "target": "Central Harlem North",
    "value": 64296.0
},
{
    "source": "Morningside Heights",
    "target": "Upper West Side",
    "value": 61455.0
},
{
    "source": "Hunters Point-Sunnyside",
    "target": "Hunters Point-Sunnyside",
    "value": 60199.0
}
```

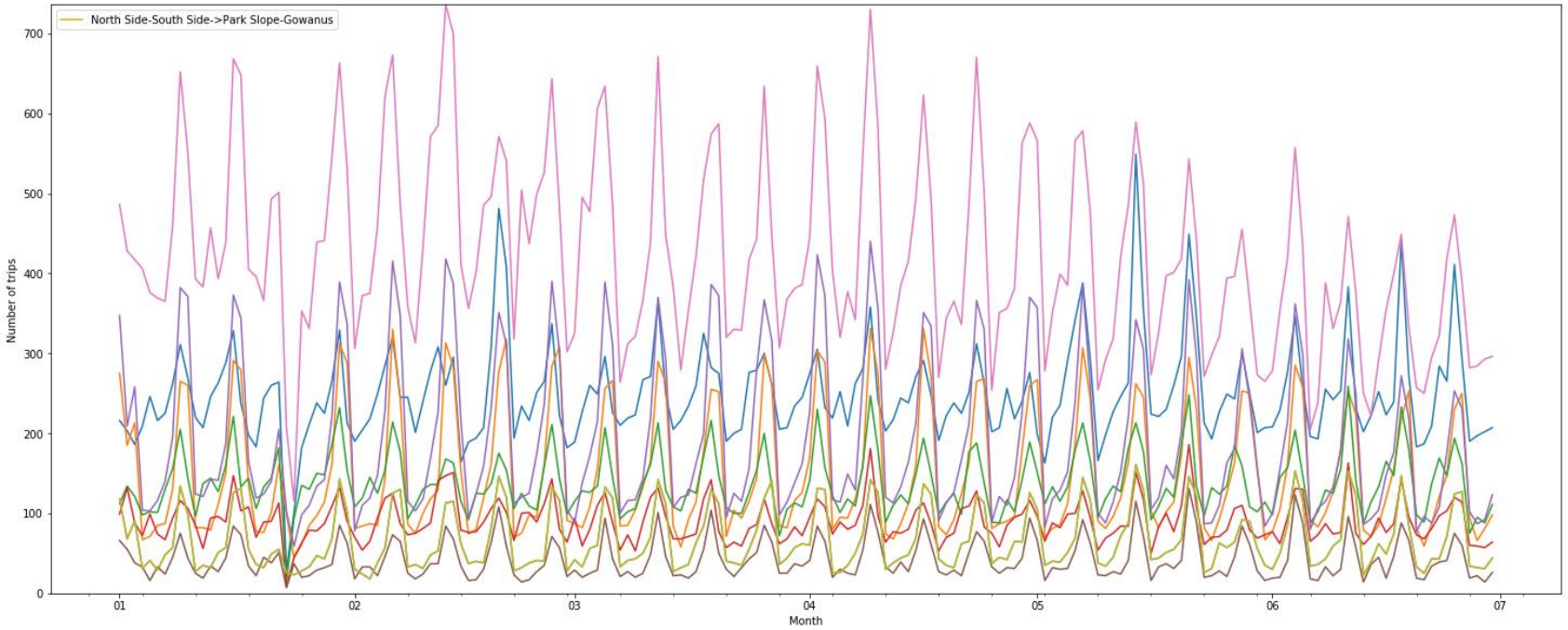
Mobility Pattern Discovery & Forecasting

- Visualization: matrix diagram



Mobility Pattern Evolution across Time

- Create time slice of transition matrix
- Visualize Pattern Evolution on time



Mobility pattern evolution across time & Forecasting

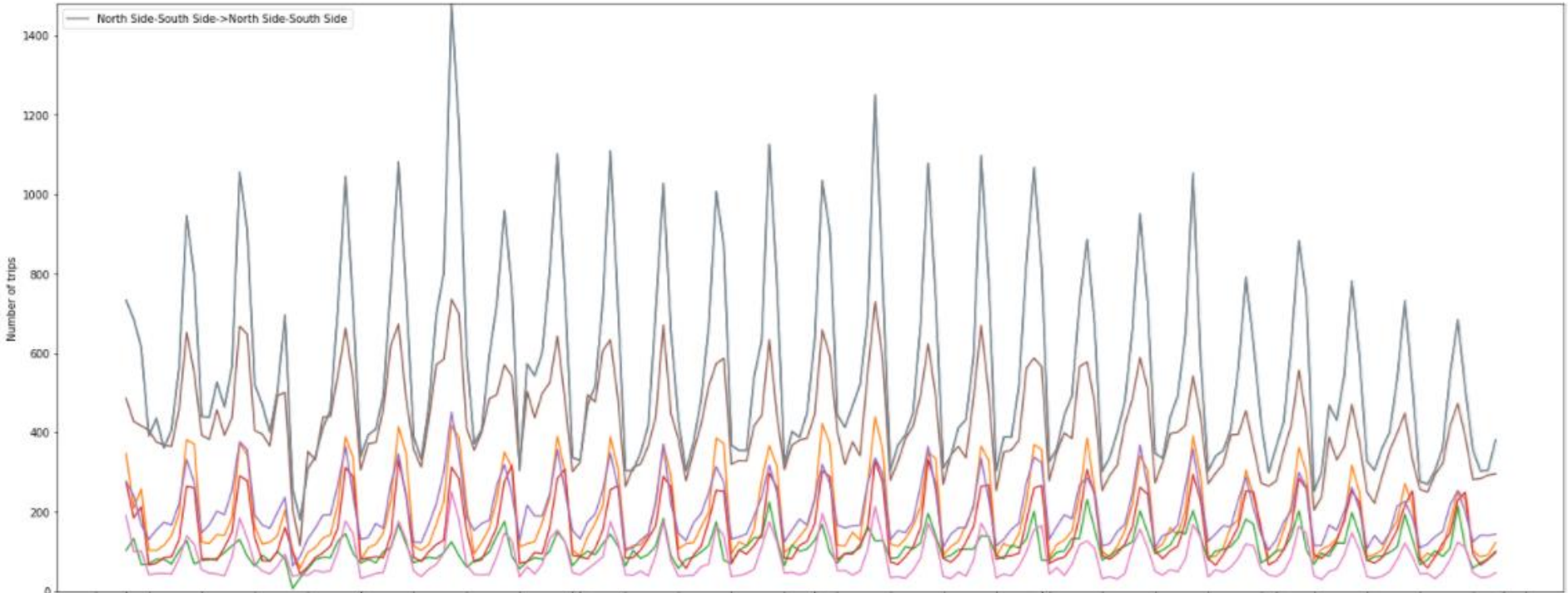
Which Transition?

From North Side-South Side To North Side-South Side Model ARIMA ARIMAX

Result

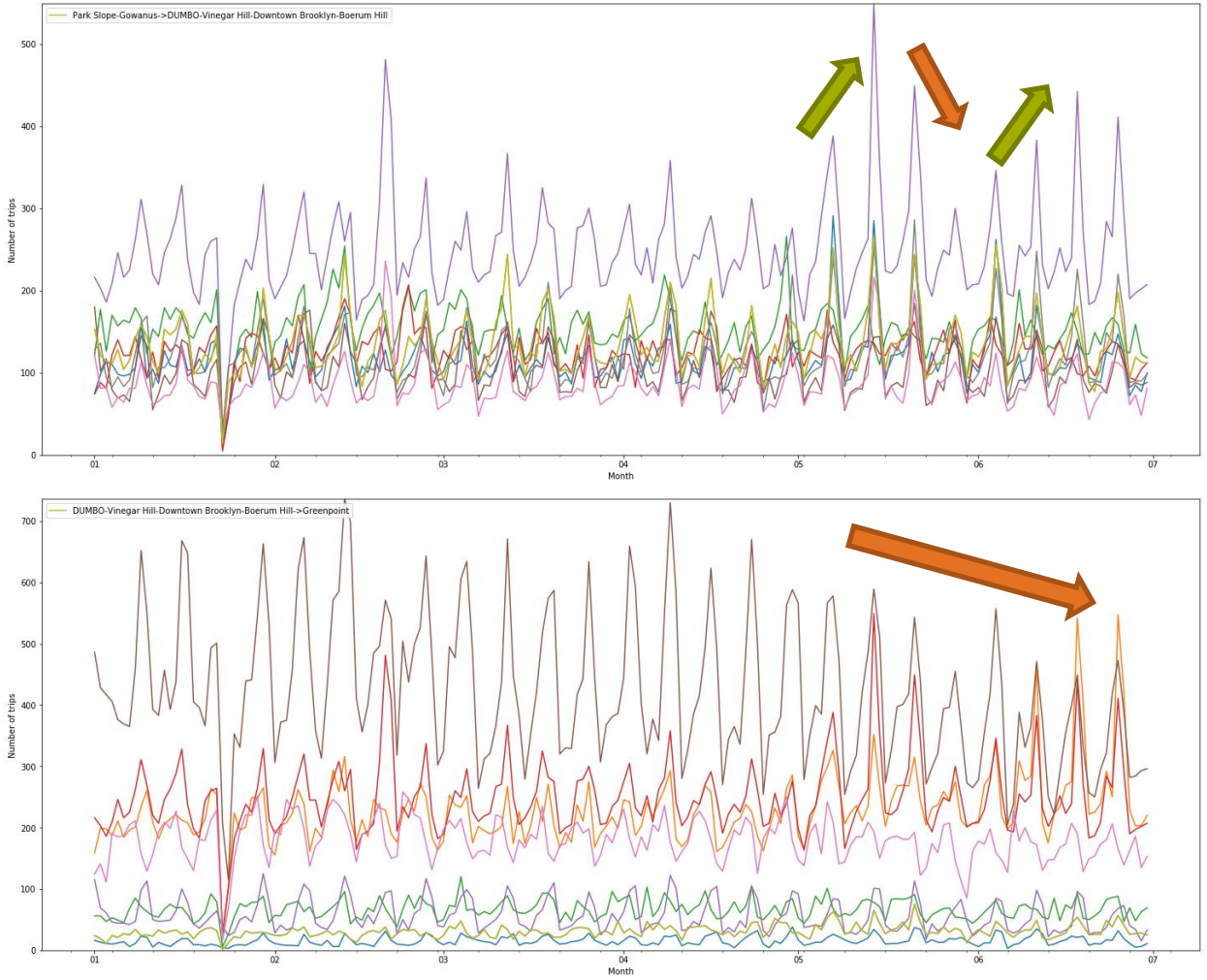
Evolution One-Step Prediction Multi-Step Prediction Parameters

The selected transition is plotted in golden. The trips from North Side-South Side to regions, having strongest connection with North Side-South Side, and trips from regions, having strongest connection with North Side-South Side, to North Side-South Side are plotted in other colors.



Mobility Pattern Evolution across Time

- Observation:
  Pattern varies on time and location

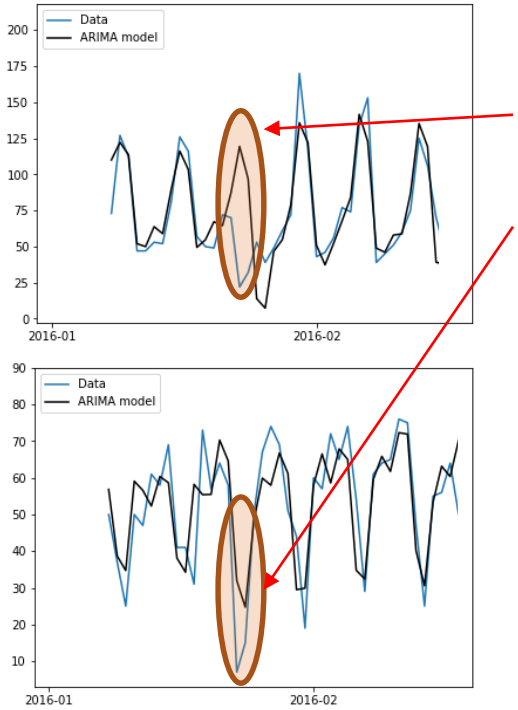


Mobility Pattern Forecasting

- Application:
- ✓ Plan where it’s better to wait tomorrow
- ✓ React fast to significant local temporal change (e.g. football match)

Mobility Pattern Forecasting

- Motivation:
- ✓ Sensor of city: event detection

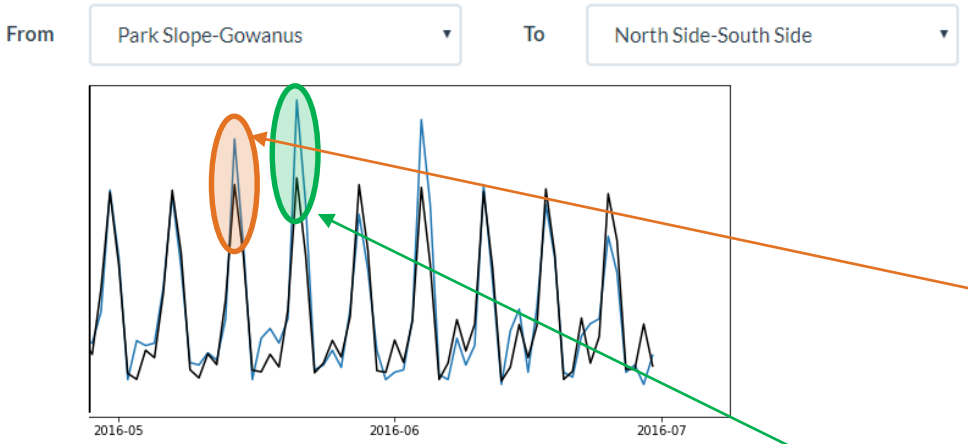


2016 Blizzard Was NYC's Biggest Snowstorm on Record, NOAA Report Finds

Jan. 22-23

Mobility Pattern Forecasting

- Motivation:
- ✓ Sensor of city: event detection



Events of Park Slope for May 2016

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
1	2	3	4	5	6	7
		Morning Moms' Group				Women's First Saturday
8	9	10	11	12	13	14
		Morning Moms' Group			Youth Group	Safe Families Training Day
15	16	17	18	19	20	21
Sunday School, Sunday Worship, Flag Football in the Park		Morning Moms' Group		Monthly Open Studios	RB Healing Service	
22	23	24	25	26	27	28
Sunday School, Sunday Worship		Morning Moms' Group			Youth Group	
29	30	31	1	2	3	4
Sunday Worship		Morning Moms' Group				Women's First Saturday

Mobility Pattern Forecasting

Model employed:

- ARIMA: Autoregressive integrated moving average

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- ARIMAX: ARIMA + Exogenous variables

$$\Delta^D y_t = \sum_{i=1}^{p} \phi_i \Delta^D y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{m=1}^{M} \beta_m Xm, t + \epsilon_t$$

$$\epsilon_t \sim N\left(0, \sigma^2\right)$$

TaxiRadar | Trip List | Mobility Pattern | Search | Submit | Contact

Mobility pattern evolution across time & Forecasting

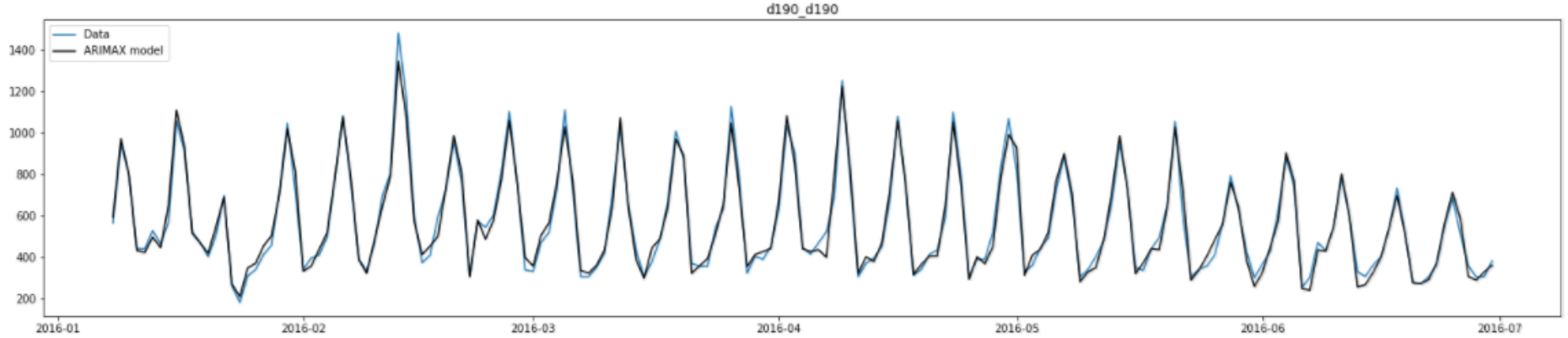
Which Transition?

From: North Side-South Side | To: North Side-South Side | Model: ARIMA | ARIMAX

Result

Evolution | One-Step Prediction | Multi-Step Prediction | Parameters

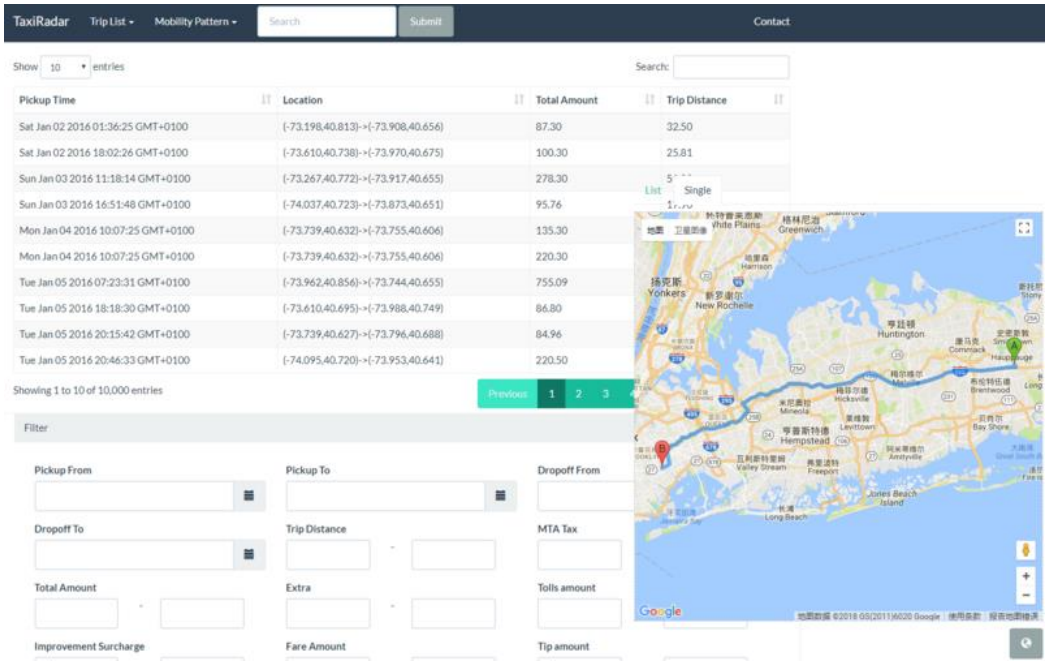
We plot rolling predictions for one day.



TaxiRadar: Online Dashboard

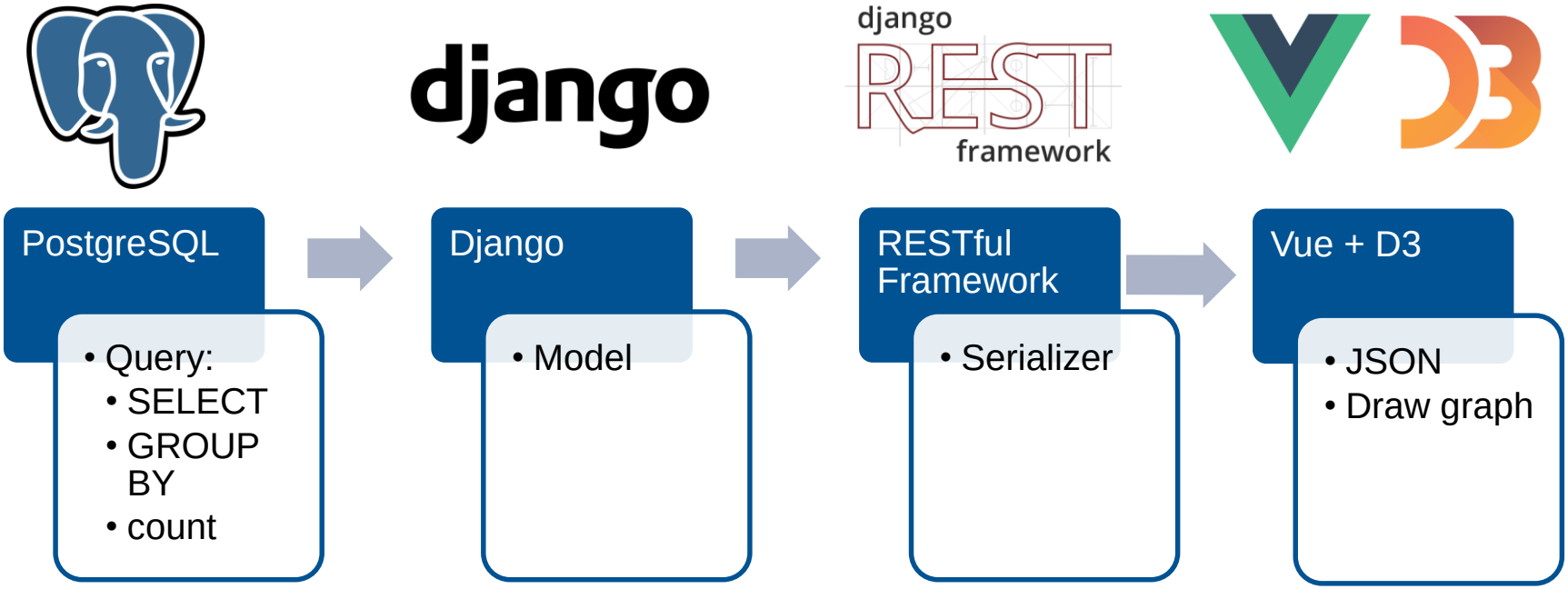
Functionality:

- Sort, filter and visualize records of trips
- Current transition matrix of mobility pattern
- Learn about our traffic forecasting.



TaxiRadar: Online Dashboard

Framework:



TaxiRadar: Online Dashboard

Tech tricks:

- Count-query Caching
- PostgreSQL Tuning
- Subquery Optimization
- TimeScaleDB
- Table Partitioning
- Columnar Storage
- Modular Front-End Development