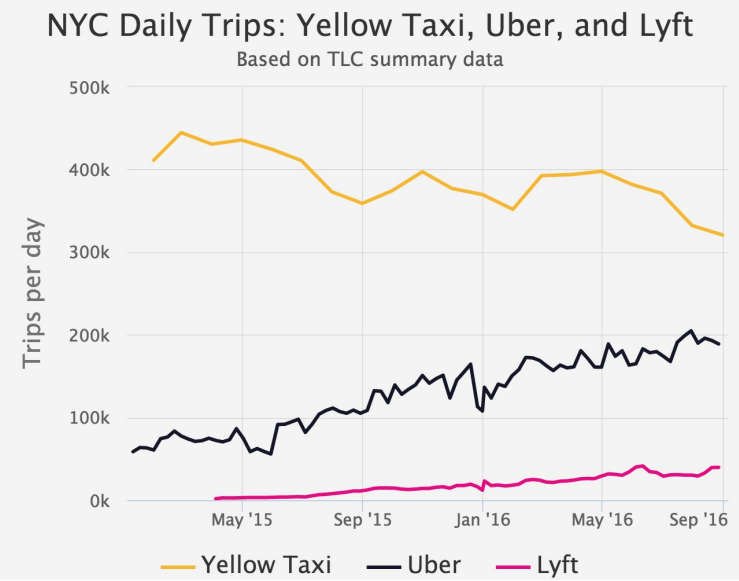
Uber’s NYC domination

Taichi Hoshino

The problem statement



Situation

- Ride-sharing is taking off in NYC
- With new entrants, how should Uber consolidate its dominance?

Complication

- 6 months of Uber data (Apr-Sep 2014) or ~5 million rides
- Only pick-up information

Question

- What is Uber's winning formula and how can it leverage this to continuing building a protective moat in NYC?

Defining Uber's winning formula

1

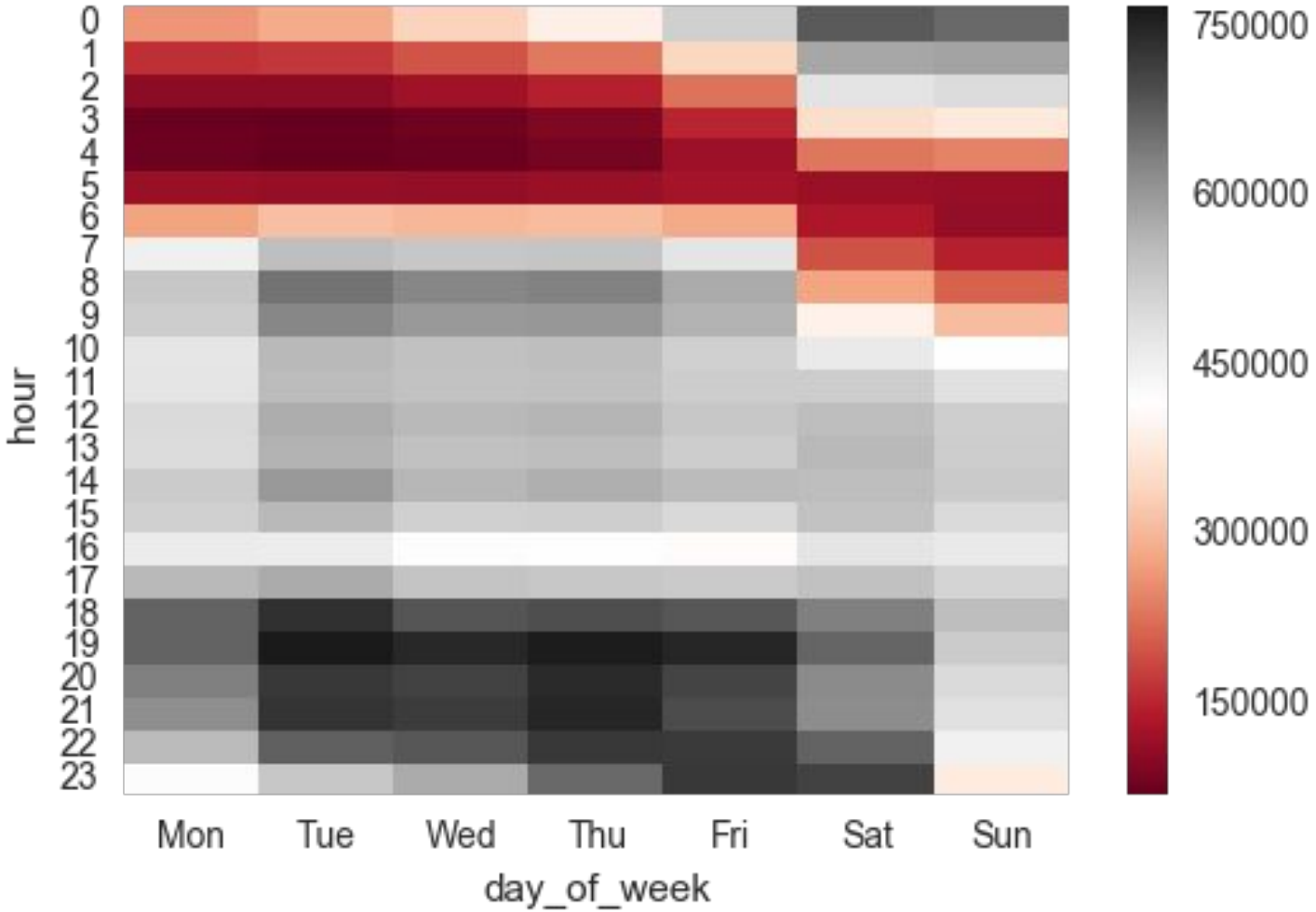
Lowest on cost

Focus of today

2

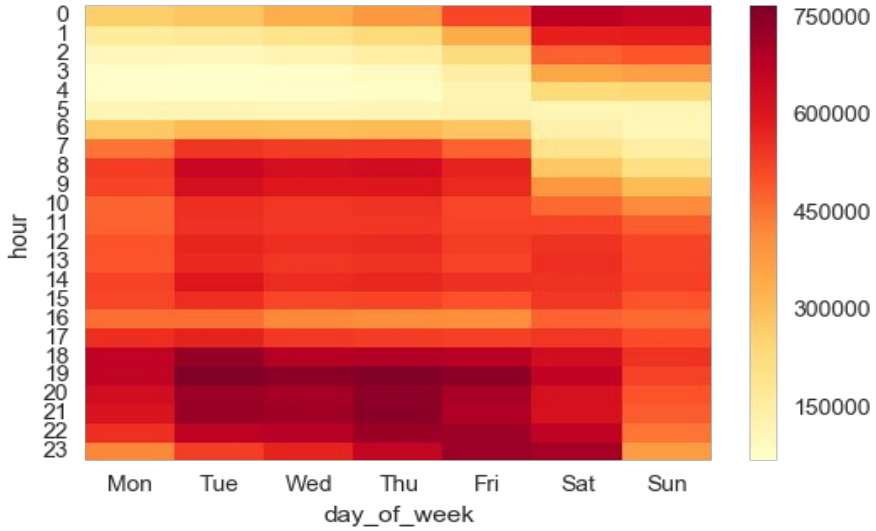
Differentiation

Taxi demand in the aggregate...

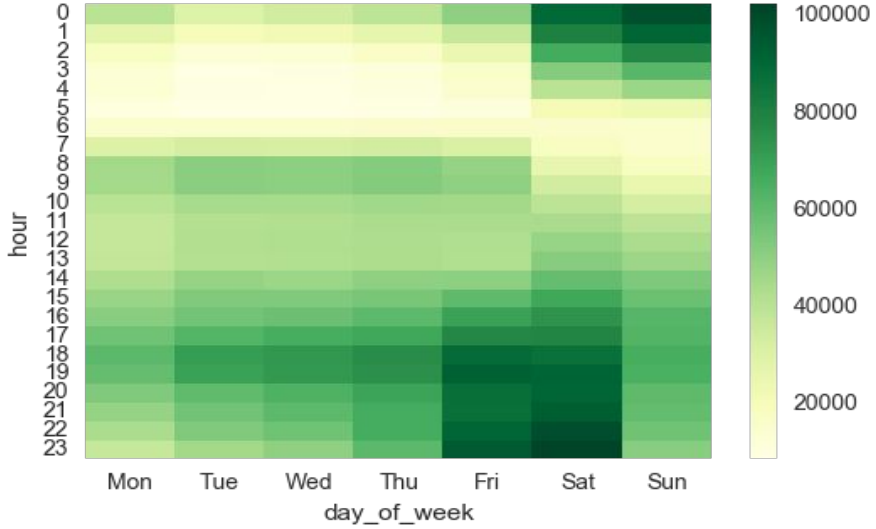


...does not differ that much by carrier

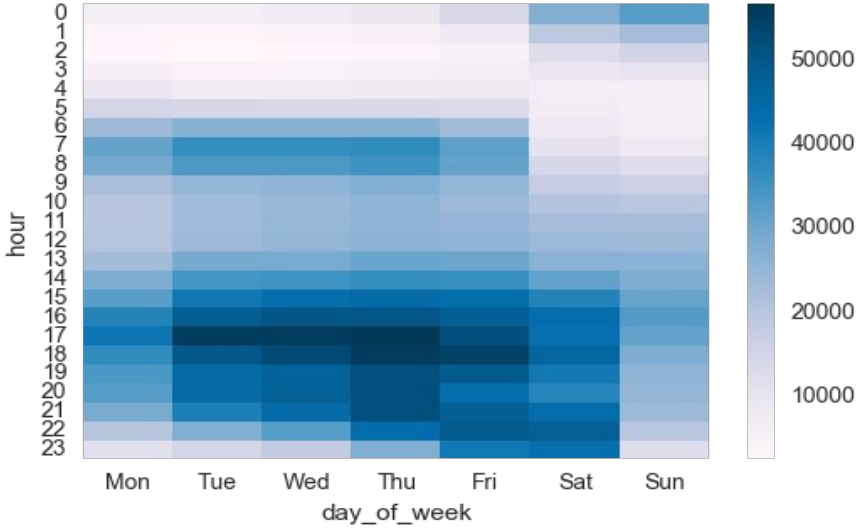
Yellow



Green

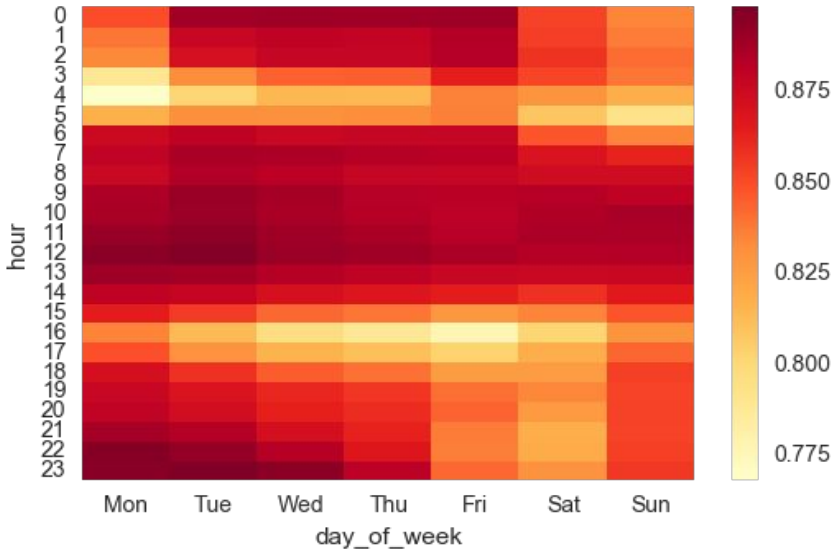


Uber

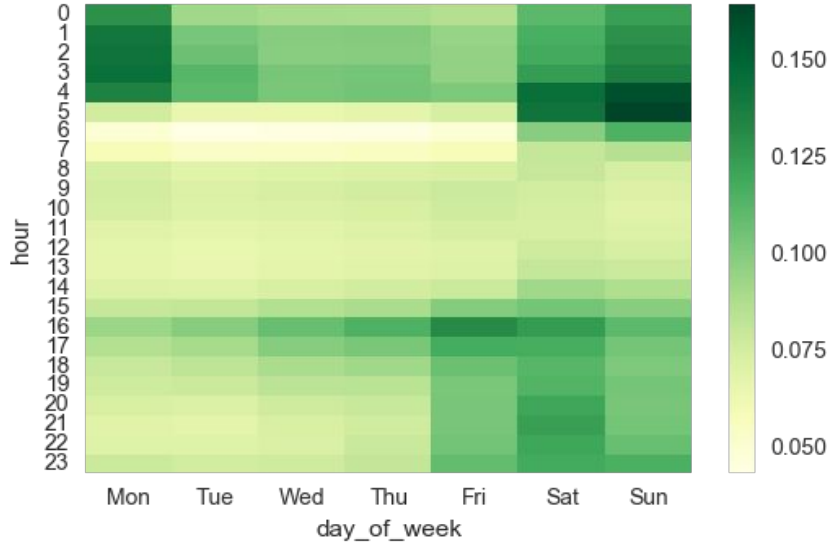


Relative market share differs materially

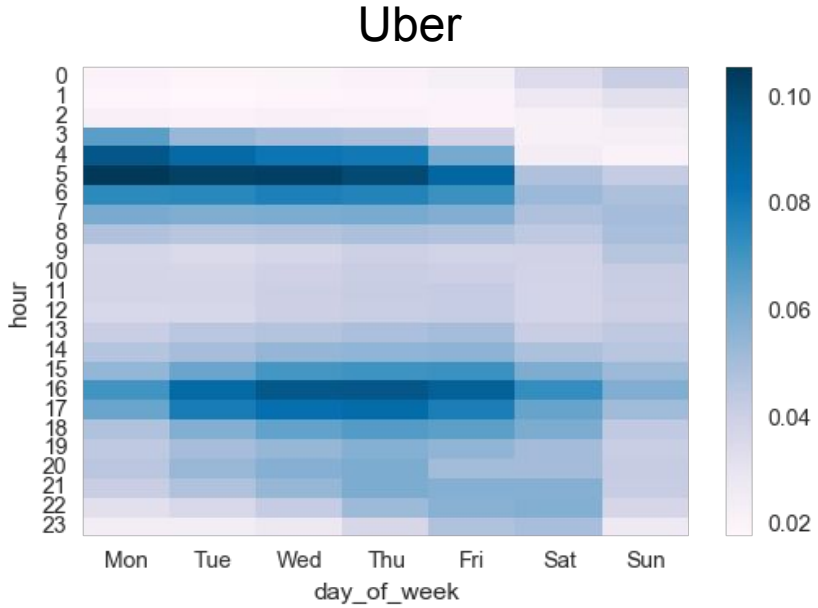
Yellow



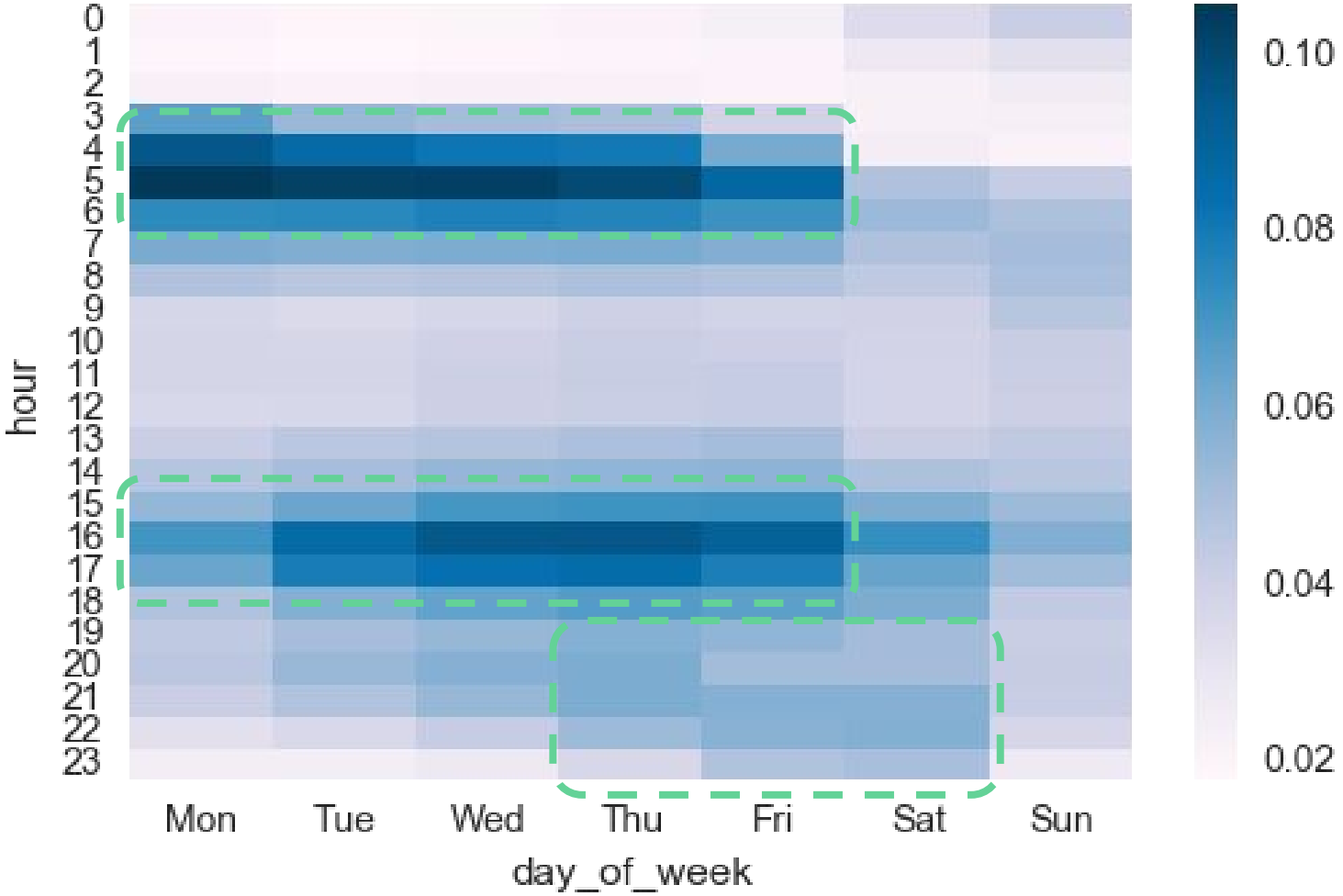
Green



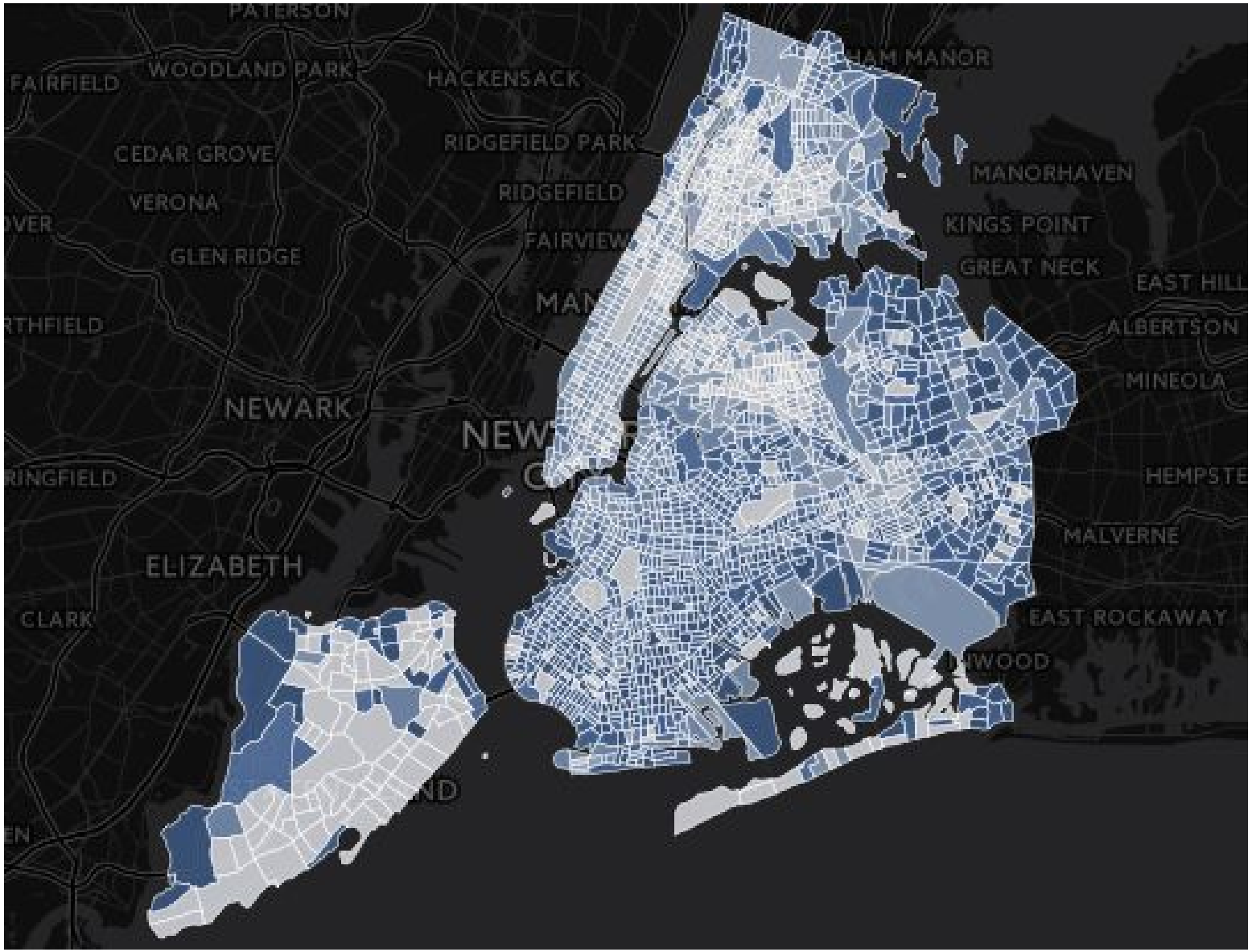
Uber



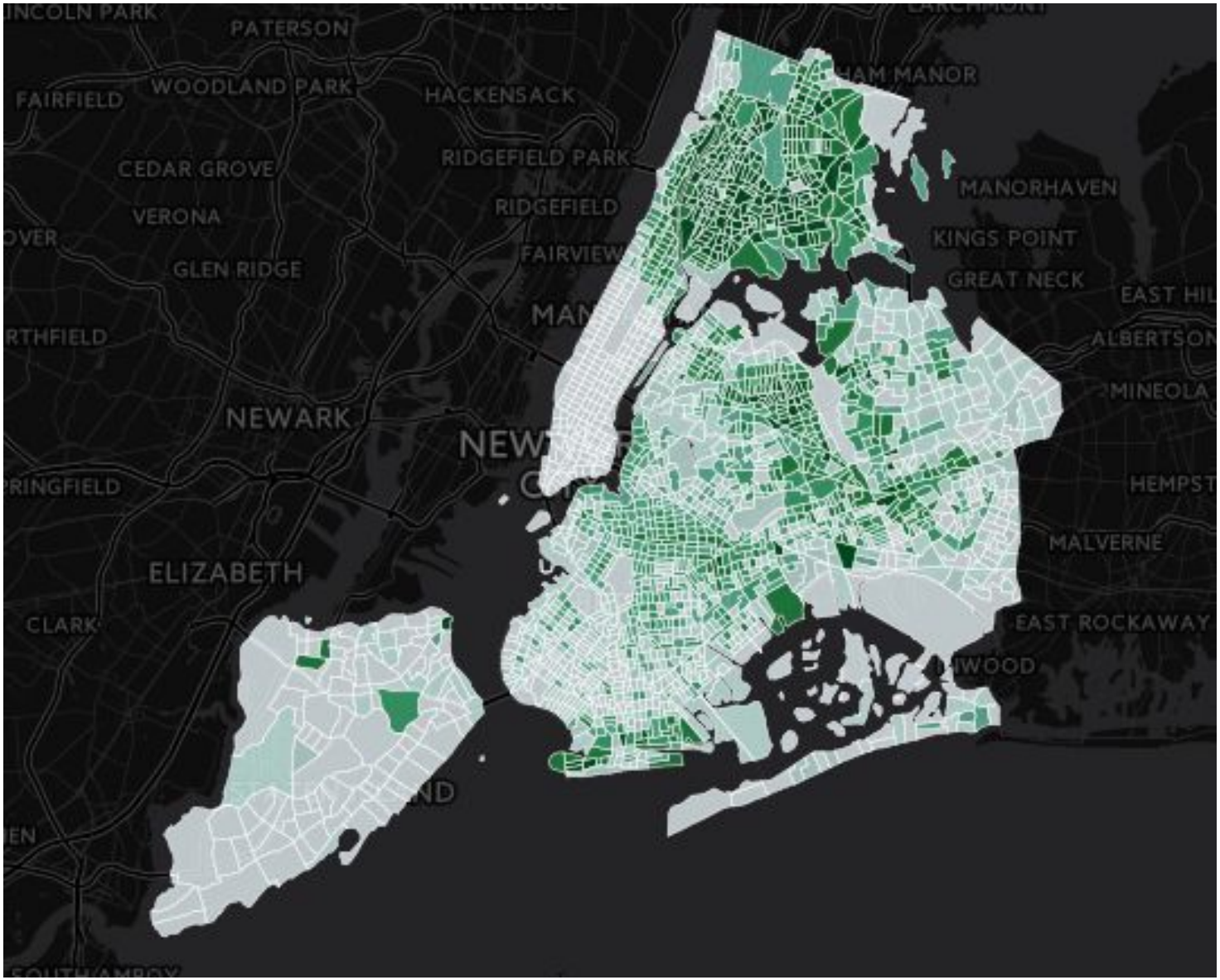
Three areas where Uber is winning



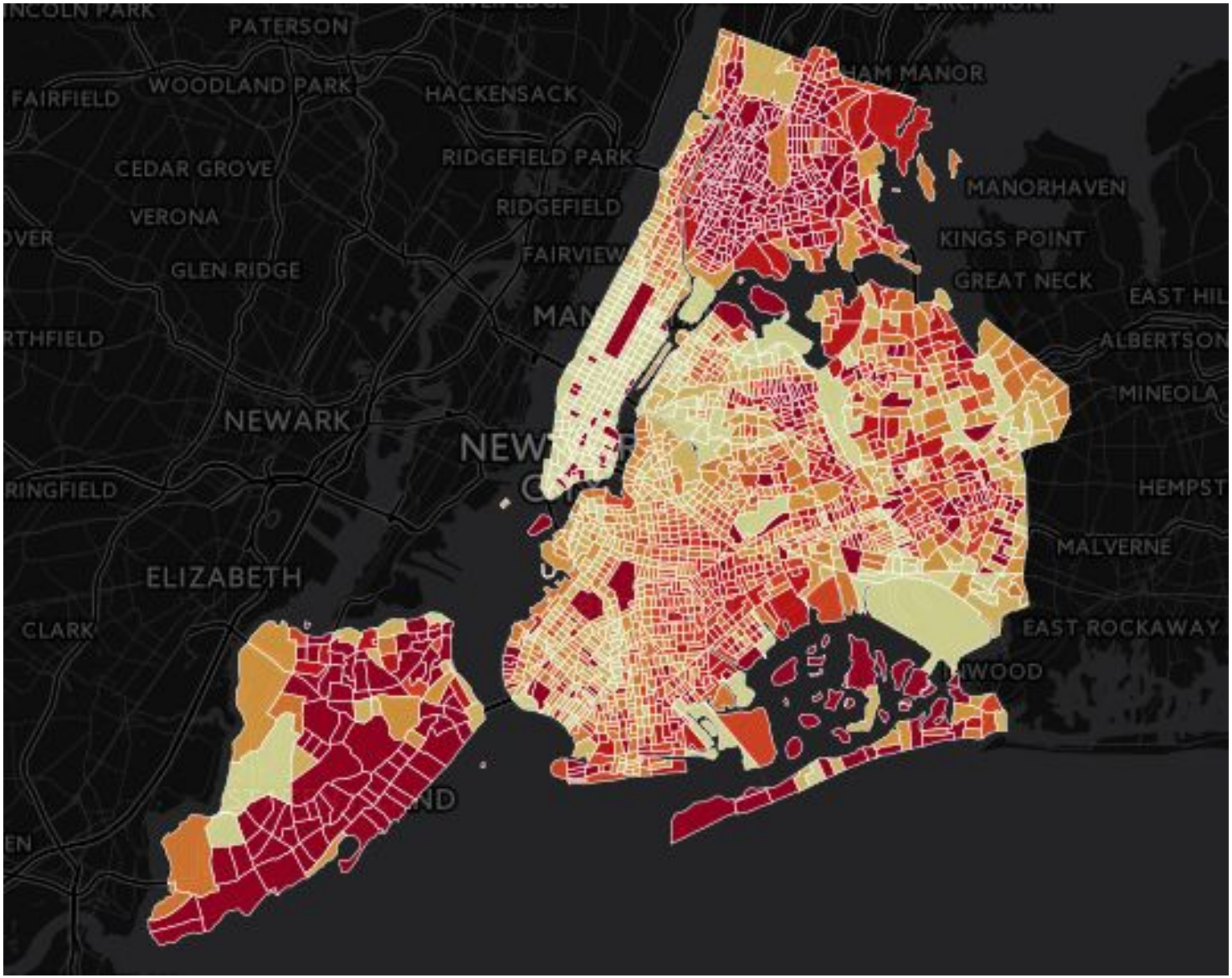
Uber share by census tract



Green car share by census tract



Yellow share by census tract



Modelling objective was twofold

Isolate features that classify census tracts where Uber had a market share of 66% of more for pick-ups?

Predict the **PEOPLE** characteristics

- ACS Demographics
  - Finances
  - Home ownership
  - Education
  - Occupation
  - Mode of transport
  - Geographic mobility

Predict the **AREA** characteristics

- Yelp businesses
- 311 noise complaints
- Taxi usage growth (5 years)
- AirBnB listings

Model selection

Models

- KNN
- LR
- SVC
- DecisionTree
- **GradientBoost**
- AdaBoost
- RandomForest

Metrics

- Accuracy
- **Precision**
- Recall
- F1
- AUC

Tuning

- init=None
- learning_rate=0.1
- loss='deviance'
- **max_depth=5**
- **max_features=2**
- max_leaf_nodes=None
- **min_samples_leaf=9**
- **min_samples_split=5**
- n_estimators=100

Feature importance

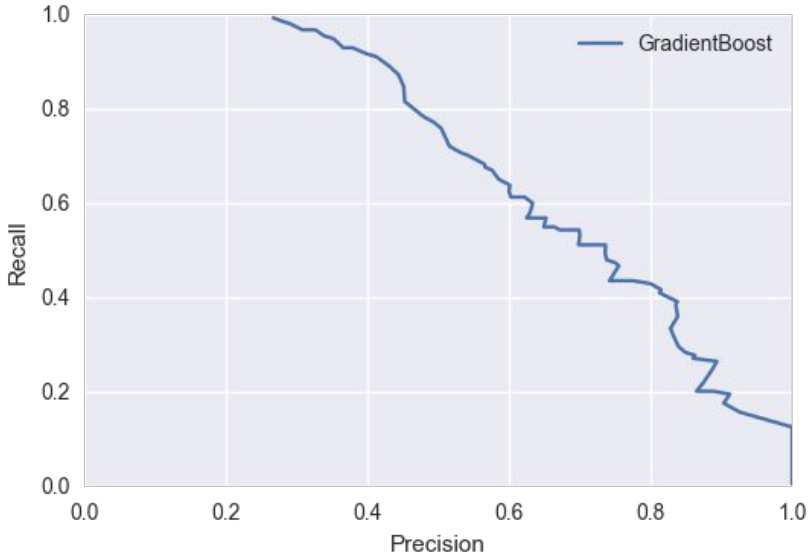
1.	**DEMO: Median monthly housing costs**	**0.1171**
2.	**AIRBNB: Listings**	**0.0952**
3.	**TLC: yellow taxi growth**	**0.0935**
4.	**DEMO: owner occupied units**	**0.0932**
5.	**311: noise complaints**	**0.0930**
6.	**DEMO: median house value**	**0.0834**
7.	**DEMO: Income per capita**	**0.0659**
8.	**YELP: review count - bars**	**0.0544**
	RANDOM	0.0520
9.	DEMO: abroad_1_year_ago	0.0506
10.	DEMO: nonfamily households	0.0505
11.	DEMO: Master's degree	0.0477
12.	DEMO: transportation public	0.0472
13.	DEMO: transportation bicyle	0.0294
14.	DEMO: renter occupied units	0.0262

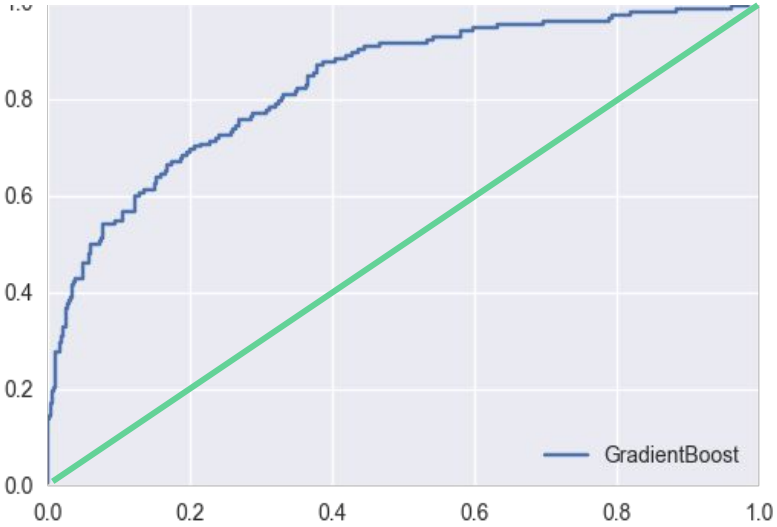
Model performance

Model: **GradientBoost**

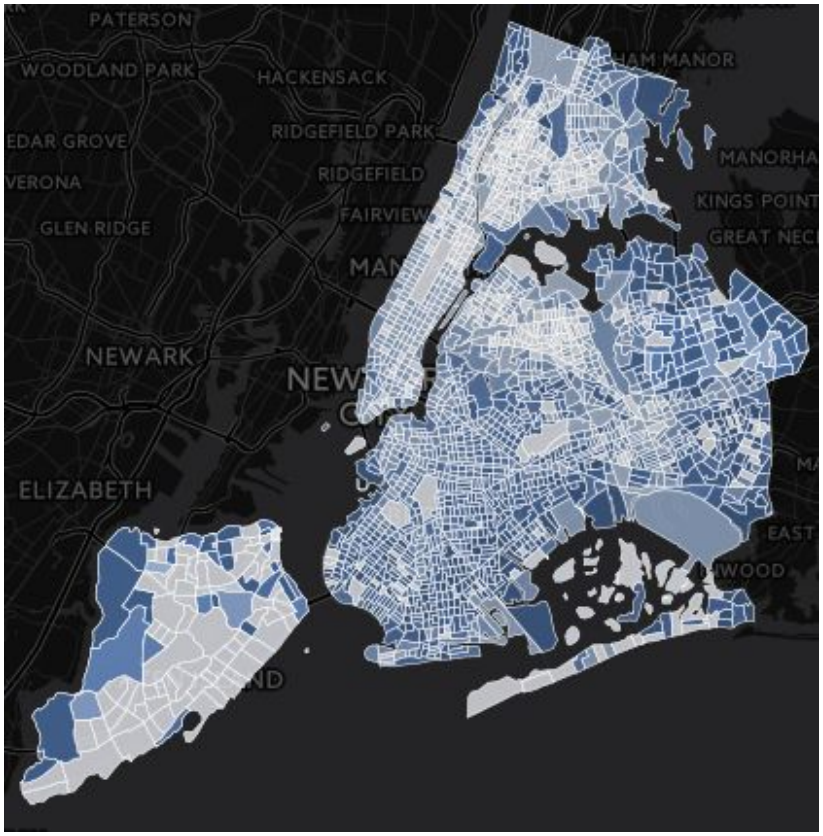
Metrics:

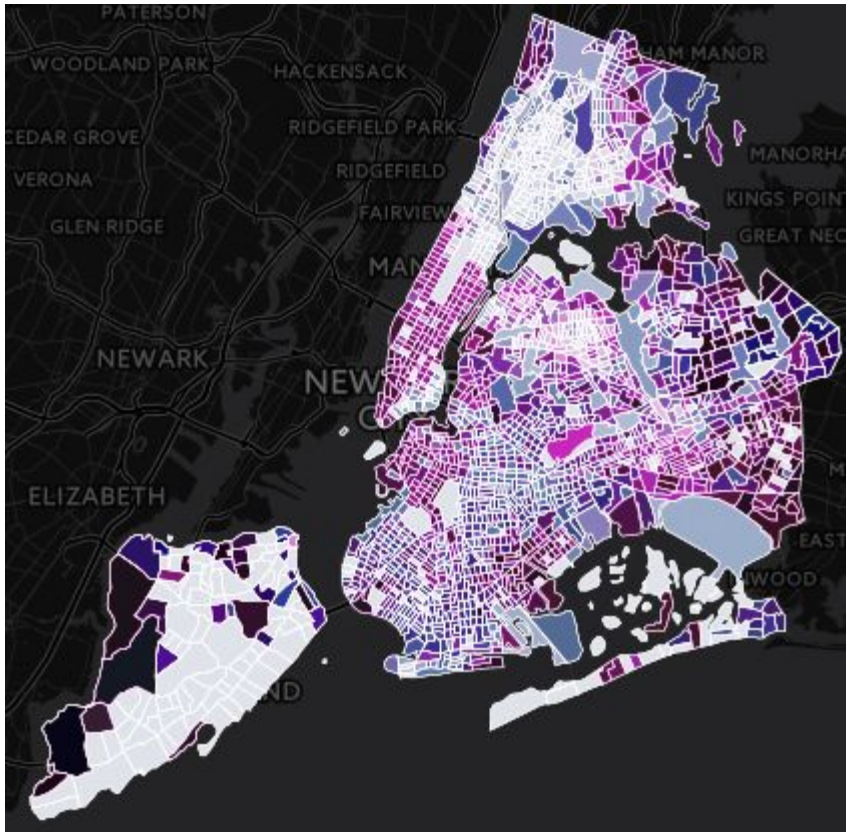
Accuracy	0.82
Precision	0.81
Recall	0.82
F1-Score	0.81
AUC	0.83



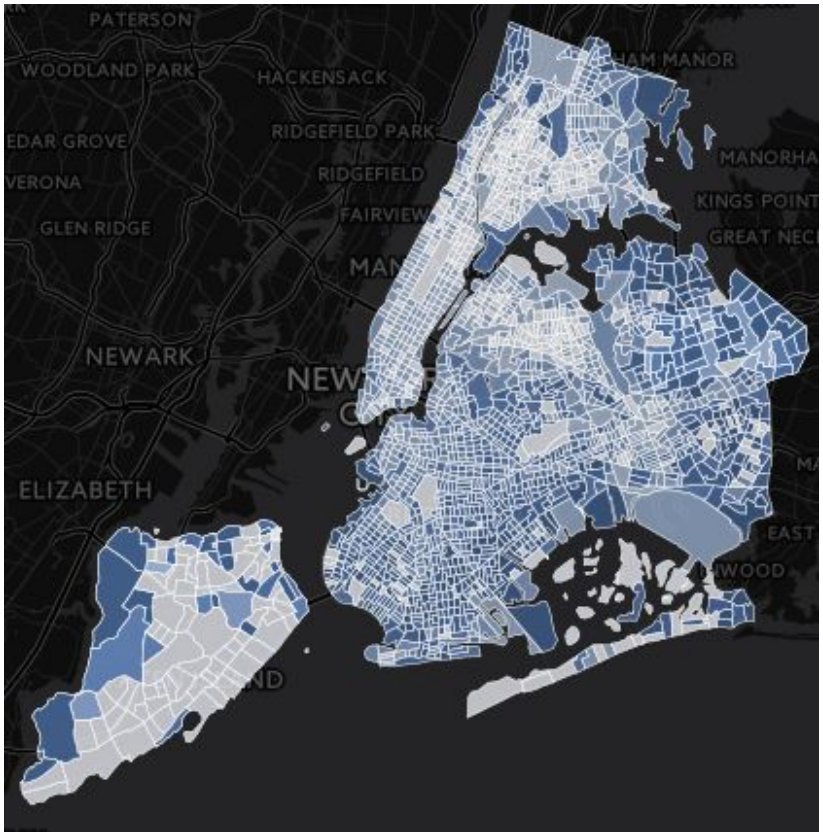


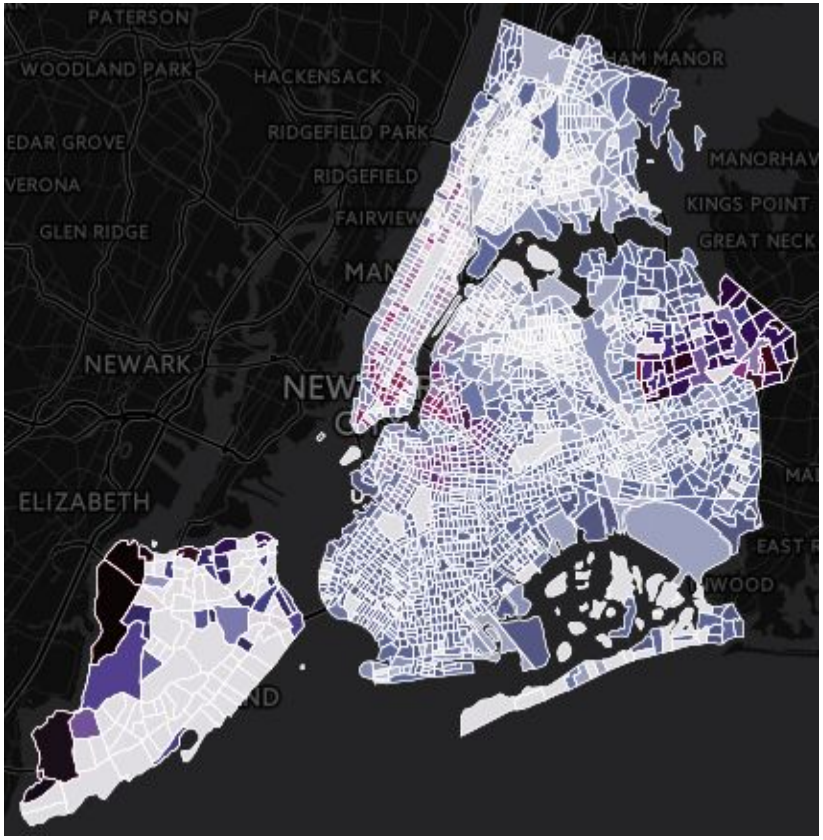
Uber share vs. median monthly housing costs



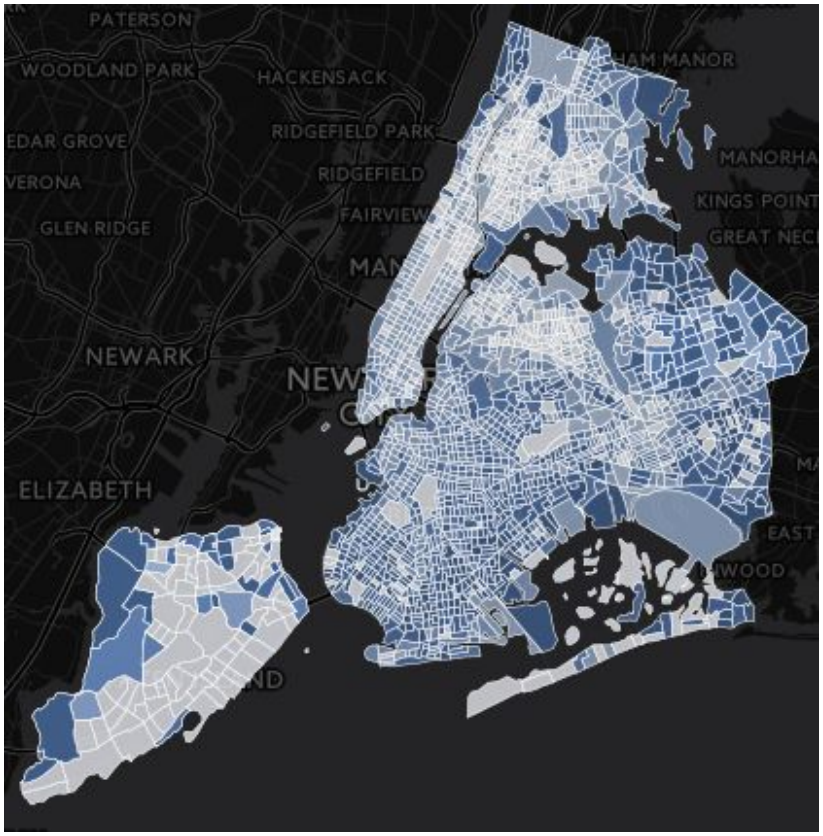


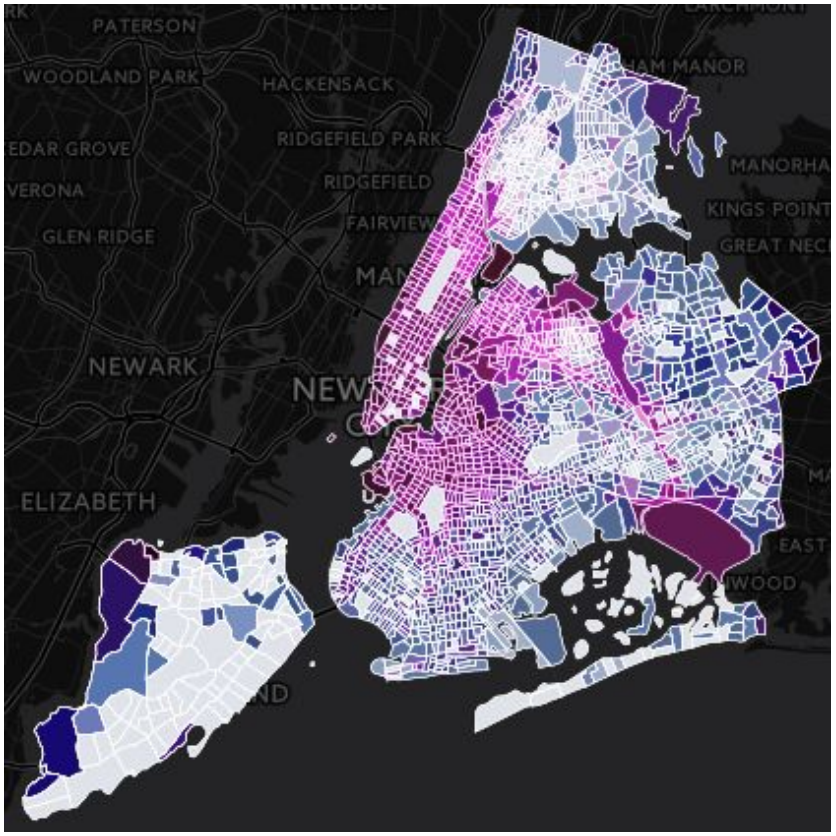
Uber share vs. Airbnb listings (#)





Uber share vs. 5 year yellow taxi growth





Implications

Marketing

- Direct spend towards people living in high potential areas with the right demographic characteristics
- Craft messages that align with one of the three differentiating value propositions

Operations

- Toggle car allocation algorithms to minimise wait times in census tracts with the highest likely pick up