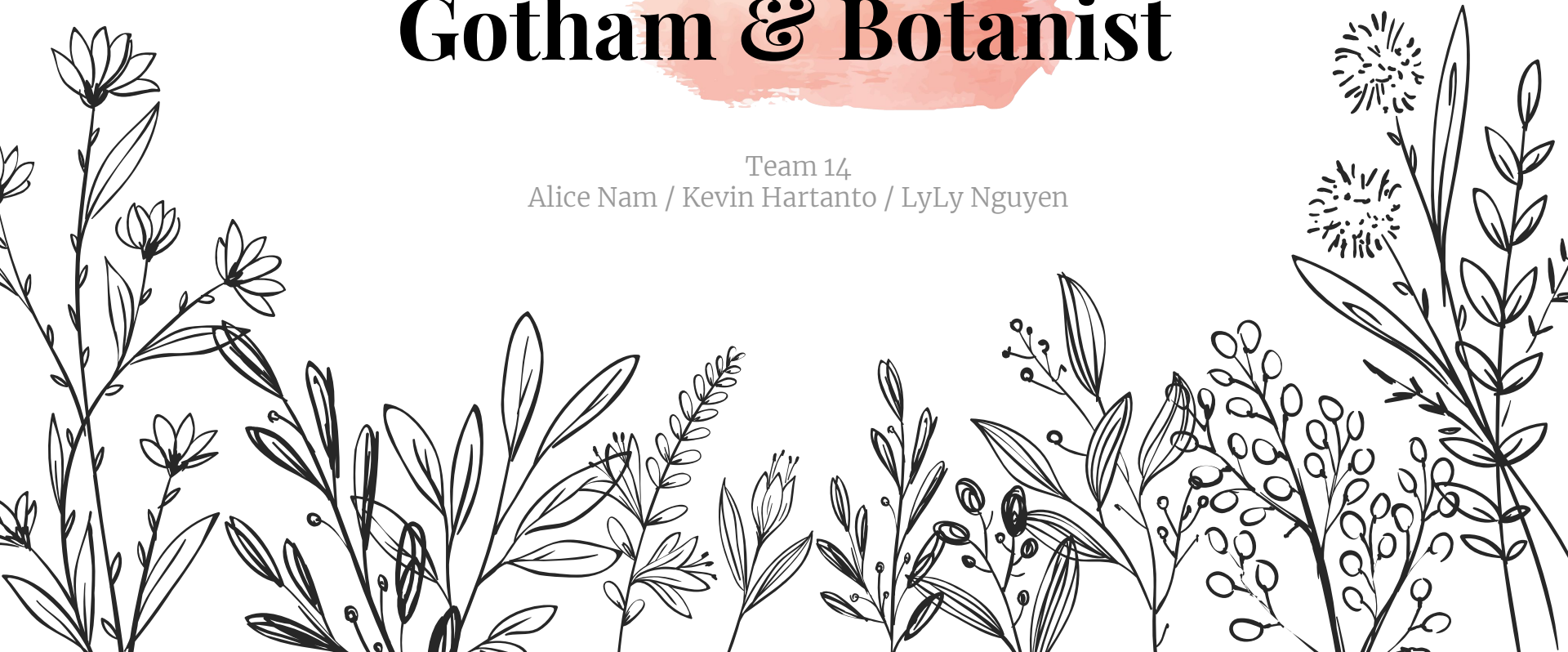
Gotham & Botanist

Team 14
Alice Nam / Kevin Hartanto / LyLy Nguyen



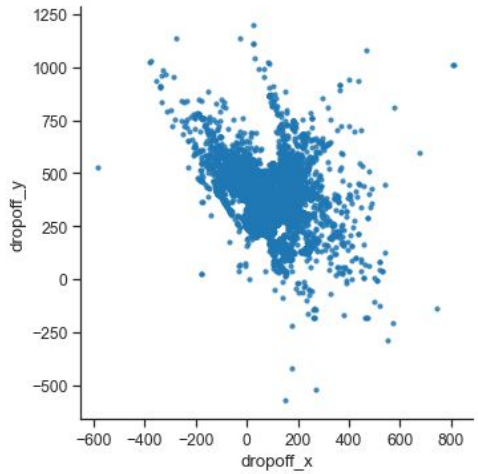
Gotham

Purpose

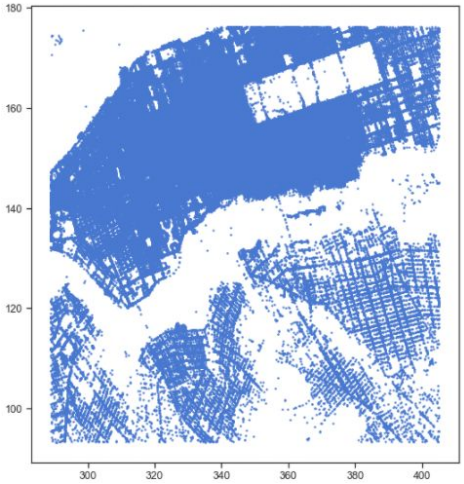
The goal of this regression problem is to train a model to predict the travel time of a cab from one city to another.

Gotham Data Processing

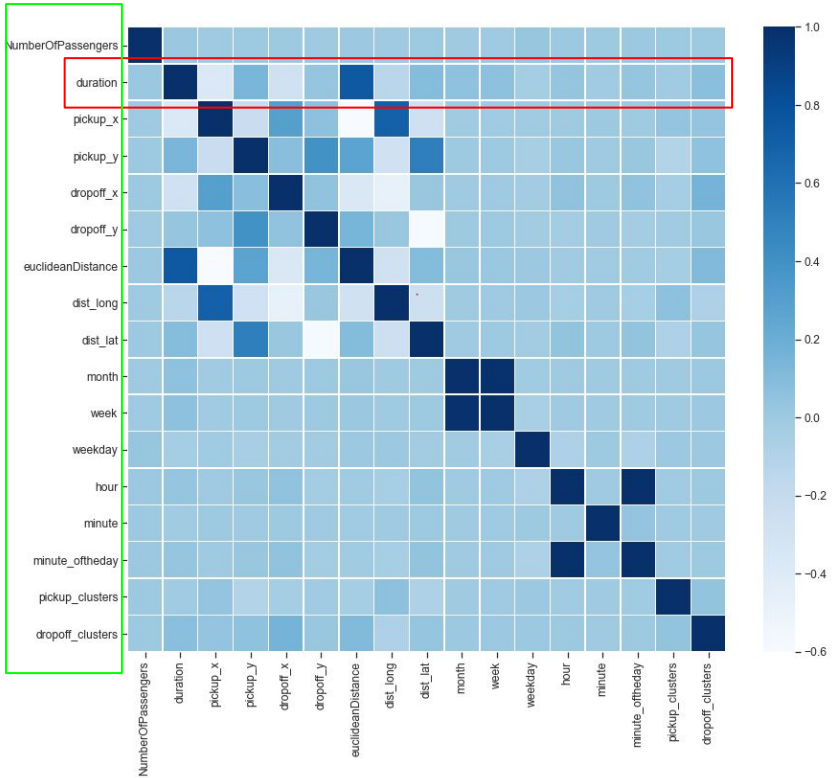
Without reduction



With reduction



Gotham Data Processing (Features)



- Drop the index, pickup_x, pickup_y, dropoff_x, dropoff_y, pickup_datetime
- Add Euclidean Distance, Distance X (dist_long) and Distance Y (dist_lat)
- Apply k-means to build grouping for the pickup and dropoff variables

Gotham Data Processing (Features)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 945458 entries, 0 to 945457
Data columns (total 13 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   NumberOfPassengers  945458 non-null  int64
 1   duration            945458 non-null  int64
 2   euclideanDistance   945458 non-null  float64
 3   dist_long           945458 non-null  float64
 4   dist_lat            945458 non-null  float64
 5   month               945458 non-null  int64
 6   week                945458 non-null  int64
 7   weekday             945458 non-null  int64
 8   hour                945458 non-null  int64
 9   minute              945458 non-null  int64
 10  minute_oftheday     945458 non-null  int64
 11  pickup_clusters     945458 non-null  int32
 12  dropoff_clusters    945458 non-null  int32
dtypes: float64(3), int32(2), int64(8)
memory usage: 86.6 MB
```

Gotham Final Model and Result

Final Model:

XG Boost

Final Result:

```
[0]     train-rmse:5.49782      valid-rmse:5.49944
[100]   train-rmse:0.35047      valid-rmse:0.36818
[200]   train-rmse:0.33391      valid-rmse:0.36281
[300]   train-rmse:0.32619      valid-rmse:0.36224
[400]   train-rmse:0.31989      valid-rmse:0.36216
[464]   train-rmse:0.31674      valid-rmse:0.36207
```

```
print('RMSE score = %1.5f, n_boost_round =%d.'%(clf.best_score,clf.best_iteration))
```

```
RMSE score = 0.36195, n_boost_round =365.
```

Botanist

Purpose

The overall goal of this project is to train a model with images of leaves and predict a label corresponding to the type and disease of the plant.

Botanist Model Breakdown

Sequential

Conv2d
: Conv2D
ReLu

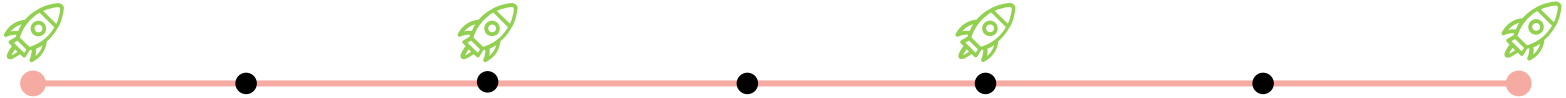
Max_pooling2d
: Maxpooling2D

Conv2d_1
: Conv2D
ReLu

Max_pooling2d_1
: Maxpooling2D

Conv2d_2
: Conv2D
ReLu

Max_pooling2d_2
: Maxpooling2D



Conv2d_3
: Conv2D
ReLu

Flatten
: Flatten

Dense
: Dense
ReLu

Dropout
: Dropout

Dense_1
: Dense
ReLu

Dropout_1
: Dropout

Dense_2
: Dense
SoftMax

83.45 %

Model Accuracy with epoch = 10

Thanks!

Any questions?