

Forecasting Visitors of Restaurants

DataScienceHK – unhackathon9

Introduction

- Project Objective:

Predict how many visitors a restaurant will have on specific day

- Data Source: Kaggle

252,108 observations of unique Restaurant-Date-Visitors combination

Dataset

- Records' date between 2016-Jan to 2017-Apr

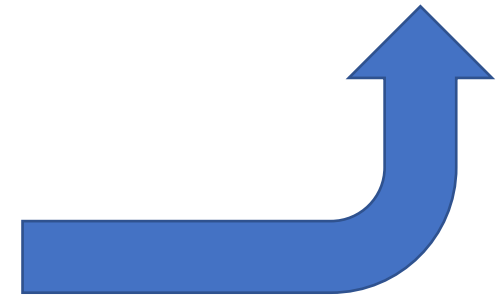
air_store_id	air_genre_name	air_area_name	latitude	longitude	visit_date	visitors	reserve_visitors	day_of_week	holiday_flg
air_e7fbee4e3cfe65c5	Izakaya	Niigata-ken Niigata-shi Gakkōchōdōri 1 Banchō	37.9164	139.036	2017-02-05	18	1633	Sunday	0
air_e7fbee4e3cfe65c5	Izakaya	Niigata-ken Niigata-shi Gakkōchōdōri 1 Banchō	37.9164	139.036	2017-02-03	33	664	Friday	0
air_e7fbee4e3cfe65c5	Izakaya	Niigata-ken Niigata-shi Gakkōchōdōri 1 Banchō	37.9164	139.036	2017-02-04	56	641	Saturday	0
air_e7fbee4e3cfe65c5	Izakaya	Niigata-ken Niigata-shi Gakkōchōdōri 1 Banchō	37.9164	139.036	2017-02-07	27	557	Tuesday	0
air_465bddfed3353b23	Dining bar	Tōkyō-to Shibuya-ku Shibuya	35.6618	139.704	2017-01-09	30	466	Monday	1
air_b23d0f519291247d	Dining bar	Fukuoka-ken Kitakyūshū-shi Ōtemachi	33.8809	130.873	2016-12-16	117	155	Friday	0
air_b23d0f519291247d	Dining bar	Fukuoka-ken Kitakyūshū-shi Ōtemachi	33.8809	130.873	2016-11-21	189	153	Monday	0
air_b23d0f519291247d	Dining bar	Fukuoka-ken Kitakyūshū-shi Ōtemachi	33.8809	130.873	2016-11-19	176	148	Saturday	0
air_a17f0778617c76e2	Italian/French	Hyōgo-ken Kōbe-shi Kumoidōri	34.6951	135.198	2017-03-16	17	142	Thursday	0
air_df554c4527a1cfe6	Izakaya	Shizuoka-ken Hamamatsu-shi Motoshirochō	34.7109	137.726	2016-12-30	205	122	Friday	1

Our Approach

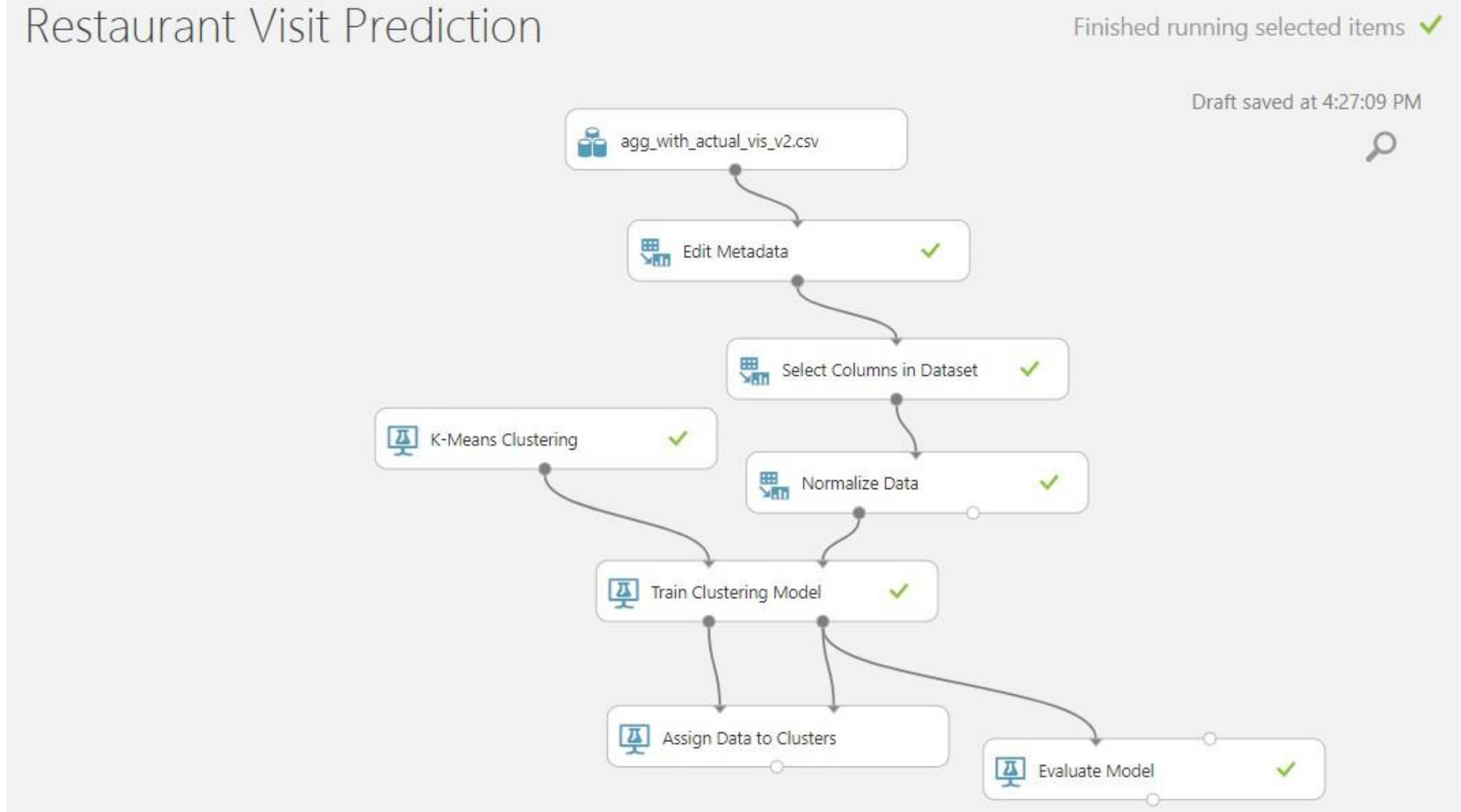
1. Transform our dataset into this table:

Restaurant	Rest Type	Rest Area	Rest Lat	Rest Long	Week Day	Holiday	Reservation	Other Features	Actual Visitors
123	Italian	Tokyo	37	139	Friday	True	34	???	????

2. Fit it in an XGBoost Decision Tree to predict actual visitors



Findings – Possible Features














Findings – Possible Features



Findings – Possible Features

Restaurant Visit Prediction > Evaluate Model > Evaluation results

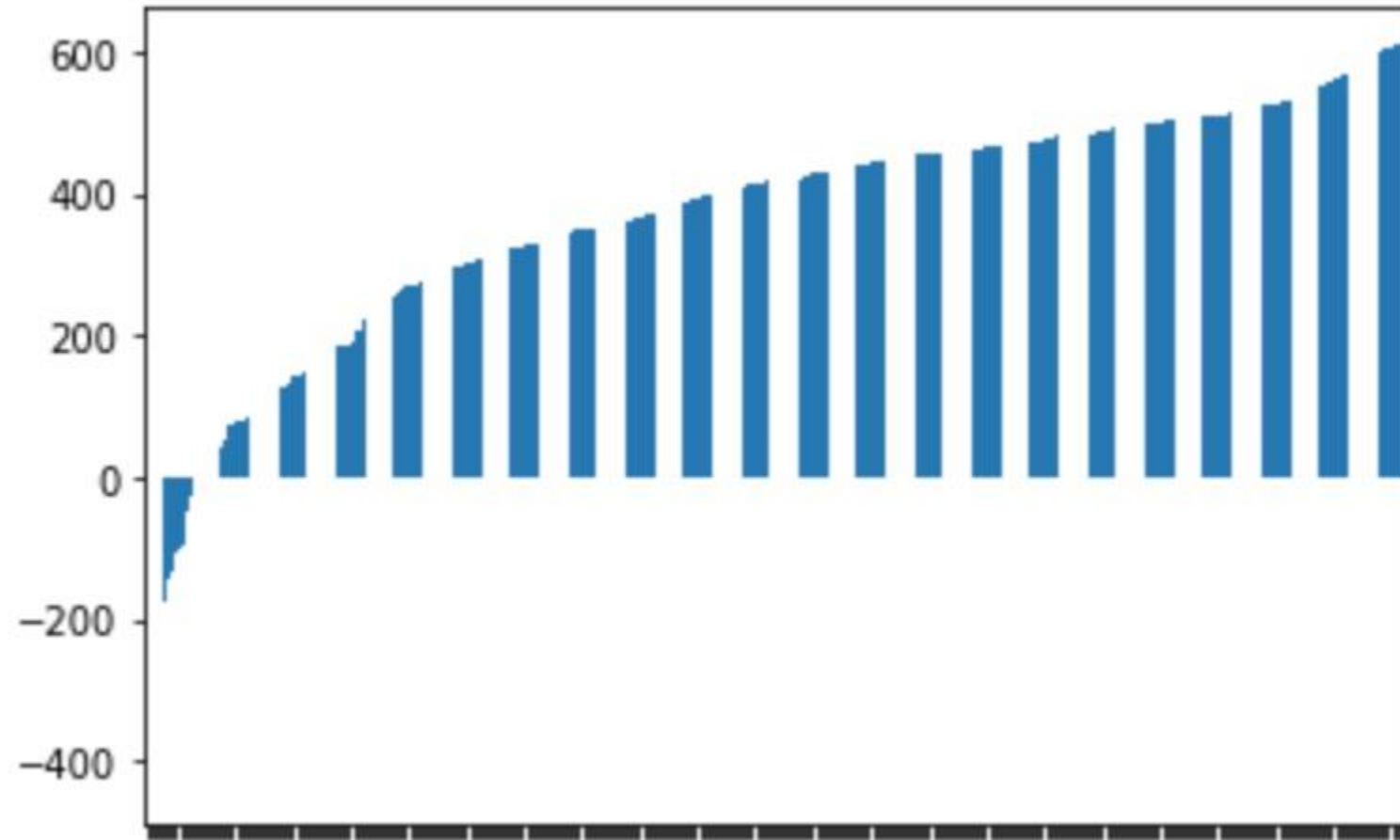
rows 3
columns 5

	Result Description	Average Distance to Cluster Center	Average Distance to Other Center	Number of Points	Maximal Distance To Cluster Center
view as  	 Combined Evaluation	  1.408206	  4.37238	  252108	  2.277119
	Evaluation For Cluster No.0	1.428947	4.375946	239053	2.277119
	Evaluation For Cluster No.1	1.028422	4.30708	13055	1.221237

Findings – Reservation as an input

	air_store_id	visit_dt_month	reserve_visitors	actual_visitors	visit_diff
0	air_00a91d42b08b08d9	2016-10-01	2	558	556
1	air_00a91d42b08b08d9	2016-10-01	2	220	218
2	air_00a91d42b08b08d9	2016-10-01	2	238	236
3	air_00a91d42b08b08d9	2016-10-01	2	32	30
4	air_00a91d42b08b08d9	2016-10-01	2	200	198

Findings – Reservation as an input

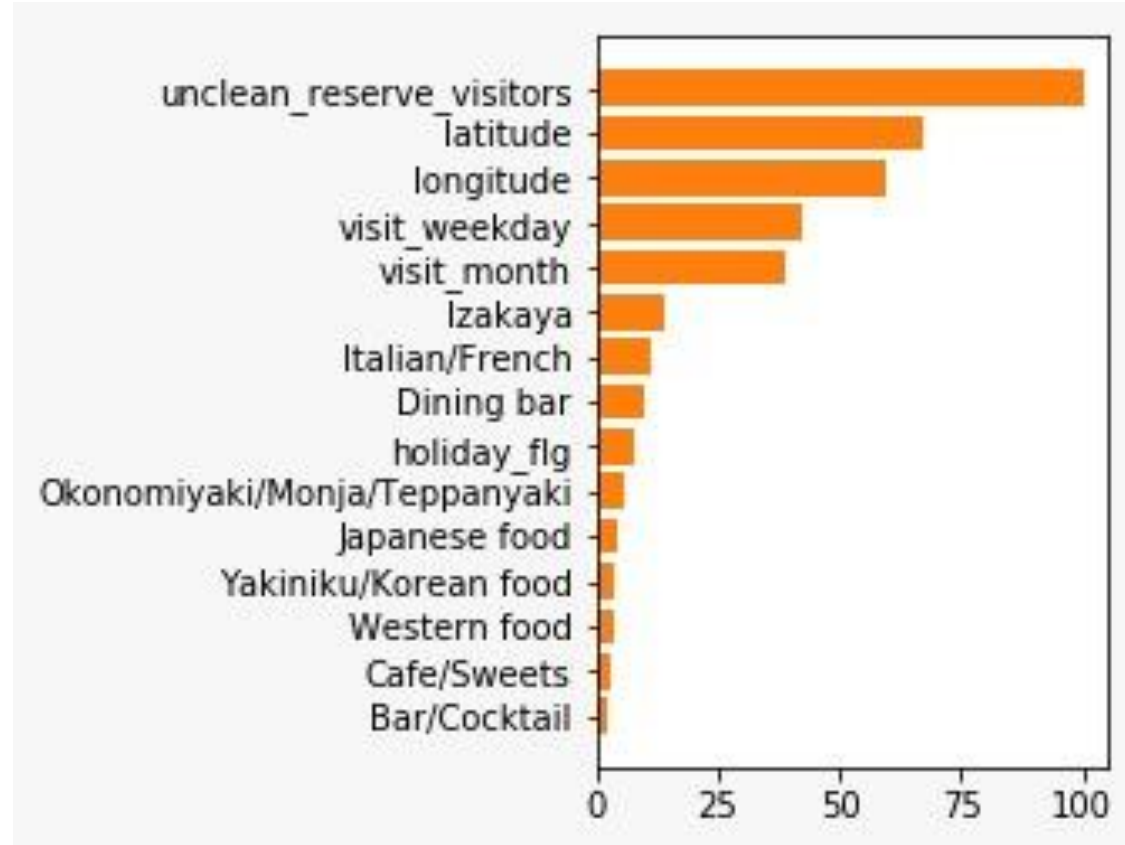


Result?

- 75 R² on training dataset
- 65 R² on testing dataset

	actual	test
15832	40.438334	38.0
11877	16.589465	35.0
11745	37.980685	42.0
23614	53.932572	68.0
6551	29.510963	28.0
6998	34.916235	37.0
28525	25.358307	24.0
20333	12.306285	66.0
40045	00.000110	00.0

Result?



Challenges

1. Further clustering restaurants did not succeed
2. Out of 252,108 observations, only 28,064 has reservations

Suggestions

Ignore reservations. Simply make prediction based on historical visitors