

CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

Introduction

⦿ Foursquare

- Built a massive dataset of locations
- Can communicate with their database through RESTful API

⦿ Nottingham

- City in England
- Estimated population of 330,000

Business Problem

- ⦿ Any link between the area of Nottingham and the type of venue that will be successful?
- ⦿ Target audiences are existing and prospective new venue owners in Nottingham

Data Acquisition

- ⦿ Combination of different datasets through Foursquare API calls
 - venue category or type
 - its location
 - how long it operated
 - how popular it was or is

Pre-processing steps

Data Acquisition

	ID	Name	Type	Latitude	Longitude	Status	StartDate	EndDate	Rating	Reviews
0	5448b19b498ec638a8a752da	200 Degrees Coffee	Coffee Shop	52.953184	-1.148888	Inactive	2014-10-23 07:43:23	2020-12-20 00:00:00	9.2	59
1	4b588c9af964a520685d28e3	Page 45	Bookstore	52.954104	-1.151368	Inactive	2010-01-21 17:19:22	2020-12-11 00:00:00	9.0	7
2	4bc9e8bcfb84c9b668dd1b3e	Aubrey's Traditional Creperie	Creperie	52.954262	-1.153241	Active	2010-04-17 16:58:36	2021-03-18 00:00:00	9.1	16
3	5377279f498ea7d99f33b627	Delilah	Deli / Bodega	52.953189	-1.146546	Active	2014-05-17 09:10:55	2021-01-21 00:00:00	8.8	9
4	54514d19498e1bb54bc03834	Malt Cross	Bar	52.953068	-1.152378	Inactive	2014-10-29 20:24:57	2020-11-18 00:00:00	8.7	11
...
242	53f8b835498e277704e0537f	The Abdication	Pub	52.998669	-1.137892	Inactive	2014-08-23 15:50:13	2020-12-16 00:00:00	NaN	1
243	516bf133e4b0c17e610ac3f7	Arno Vale Recreation Ground	Park	52.994394	-1.124791	Inactive	2013-04-15 12:23:15	2020-12-12 00:00:00	NaN	1
244	57f6e6fb498eacf298264432	A.C.E.S. Ltd	Other Repair Shop	52.991185	-1.131628	Active	2016-10-07 00:06:19	2021-01-07 00:00:00	NaN	0
245	4e61fbd862e13e3bce6efa50	Arno Vale Play Park	Playground	52.993740	-1.123585	Active	2011-09-03 10:05:12	2021-03-28 00:00:00	NaN	1
246	5b70445d4a1cc0002c360d01	Febuary Gardens	Bar	52.994373	-1.123004	Active	2018-08-12 14:29:49	2021-02-02 00:00:00	NaN	0

247 rows × 10 columns

Pre-processing steps

Data Cleaning

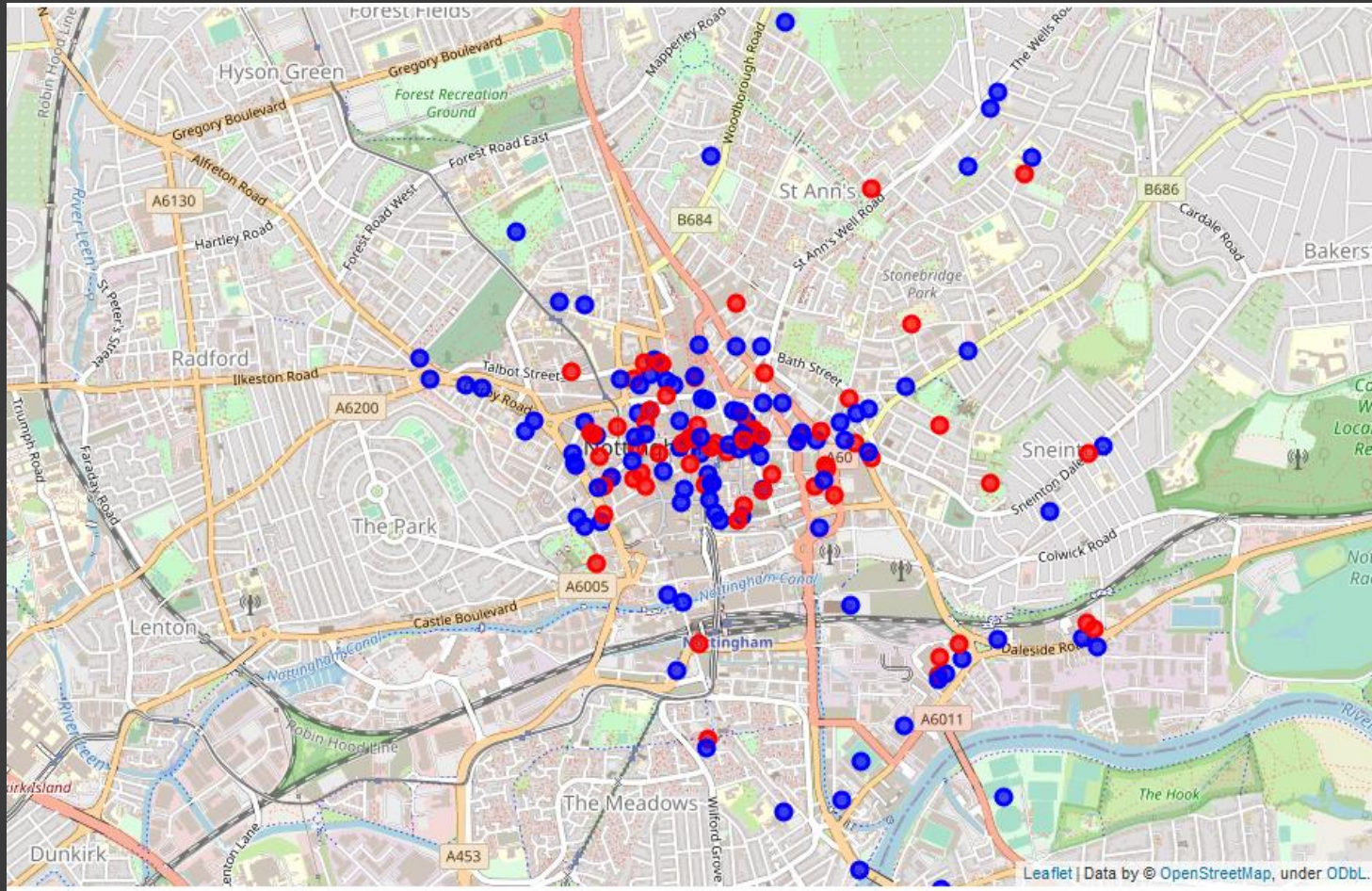
	Name	Type	Latitude	Longitude	DaysOperated	Reviews	Rating	Status
0	200 Degrees Coffee	Coffee Shop	52.953184	-1.148888	2249	59	9.2	Inactive
1	Page 45	Bookstore	52.954104	-1.151368	3976	7	9.0	Inactive
2	Aubrey's Traditional Creperie	Creperie	52.954262	-1.153241	3987	16	9.1	Active
3	Delilah	Deli / Bodega	52.953189	-1.146546	2440	9	8.8	Active
4	Malt Cross	Bar	52.953068	-1.152378	2211	11	8.7	Inactive
...
242	The Abdication	Pub	52.998669	-1.137892	2306	1	NaN	Inactive
243	Arno Vale Recreation Ground	Park	52.994394	-1.124791	2797	1	NaN	Inactive
244	A.C.E.S. Ltd	Other Repair Shop	52.991185	-1.131628	1552	0	NaN	Active
245	Arno Vale Play Park	Playground	52.993740	-1.123585	3493	1	NaN	Active
246	Febuary Gardens	Bar	52.994373	-1.123004	904	0	NaN	Active

247 rows × 8 columns

Exploratory Data Analysis

- ⦿ All venues mapping
 - All active and inactive venues
- ⦿ Food venues analysis
 - Aggregation of different venue categories
- ⦿ Relationship between parameters
 - Reviews vs. Days of Operation
 - Ratings vs. Reviews
 - Ratings vs. Status vs. Category

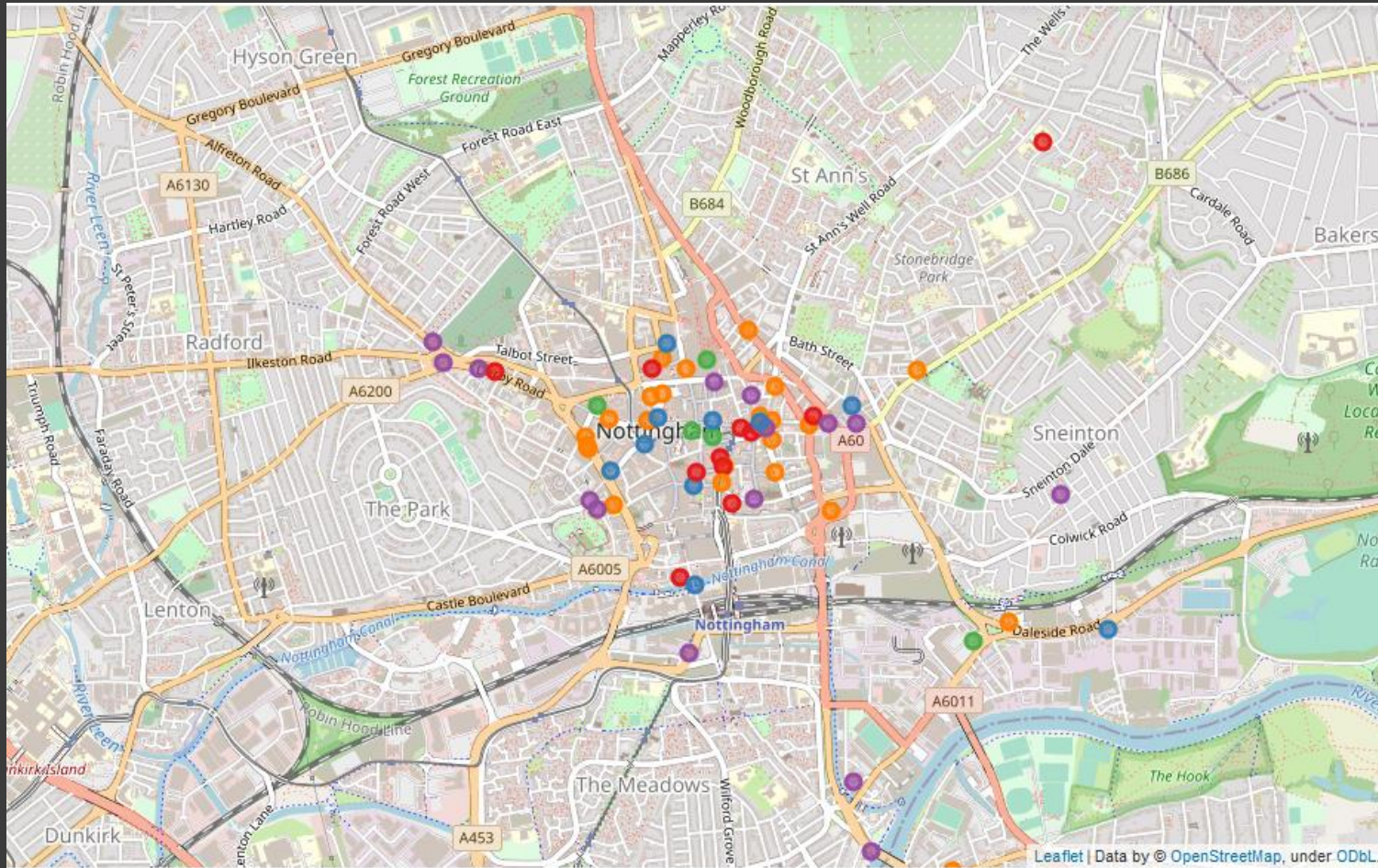
All venues mapping



ALL VENUES SHOWN ON THE MAP - ACTIVE IN BLUE, INACTIVE IN RED

Exploratory Data Analysis

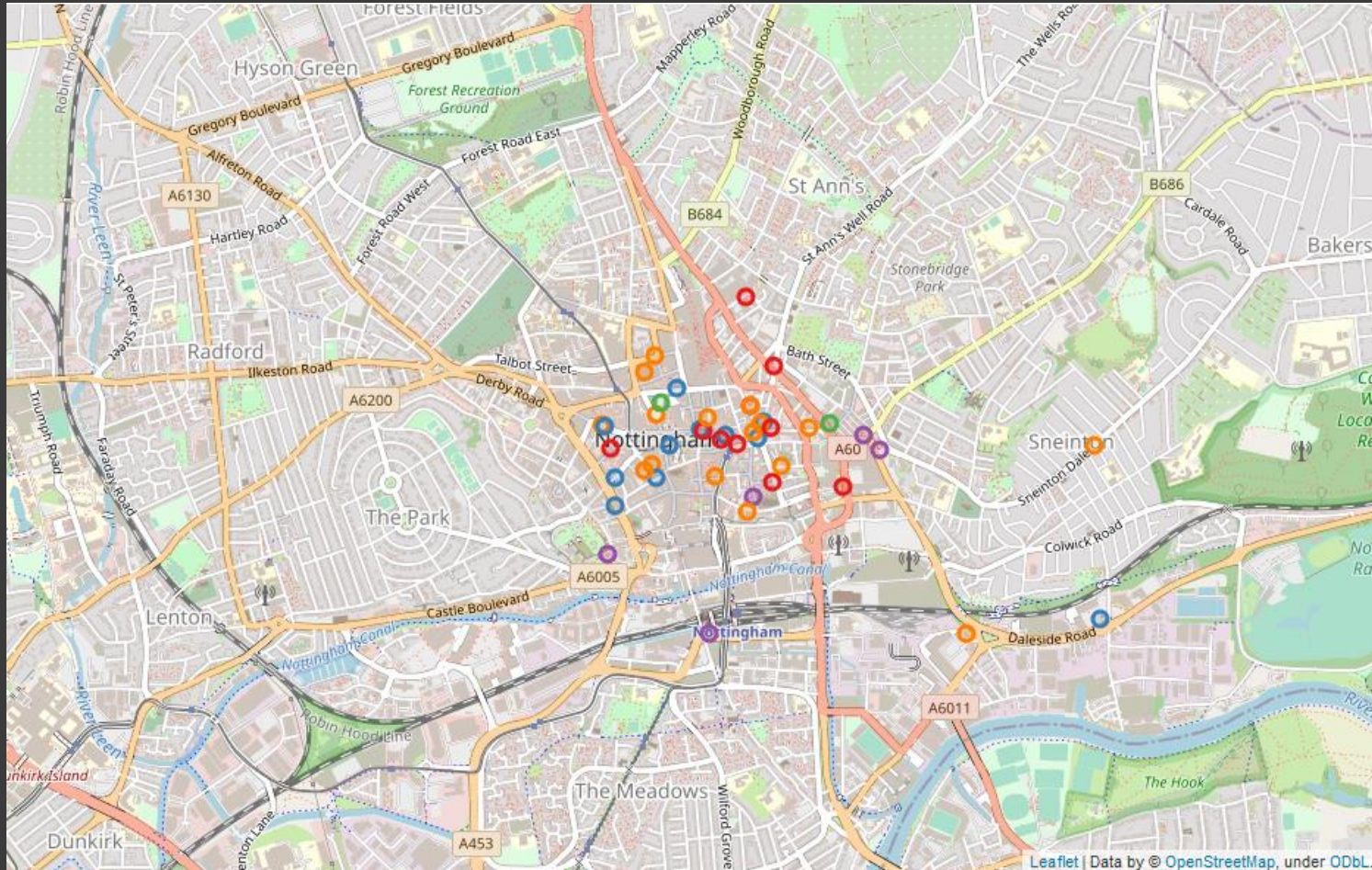
Food venues analysis



ACTIVE FOOD VENUES: RED - BAR , ORANGE - RESTAURANT, BLUE - CAFE , GREEN - DELI , PURPLE – PUB

Exploratory Data Analysis

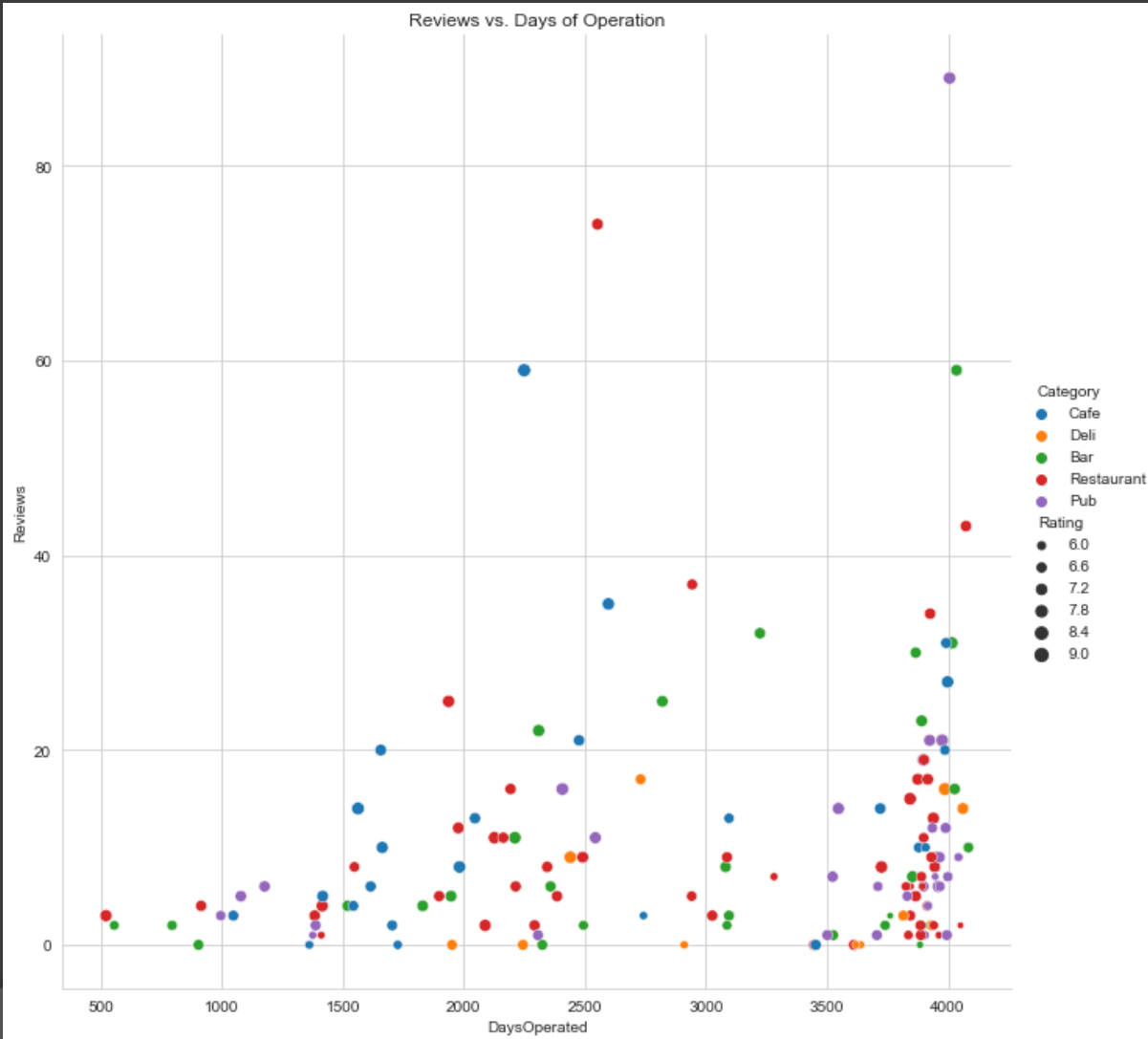
Food venues analysis



INACTIVE FOOD VENUES: RED - BAR , ORANGE - RESTAURANT, BLUE - CAFE , GREEN - DELI , PURPLE – PUB

Exploratory Data Analysis

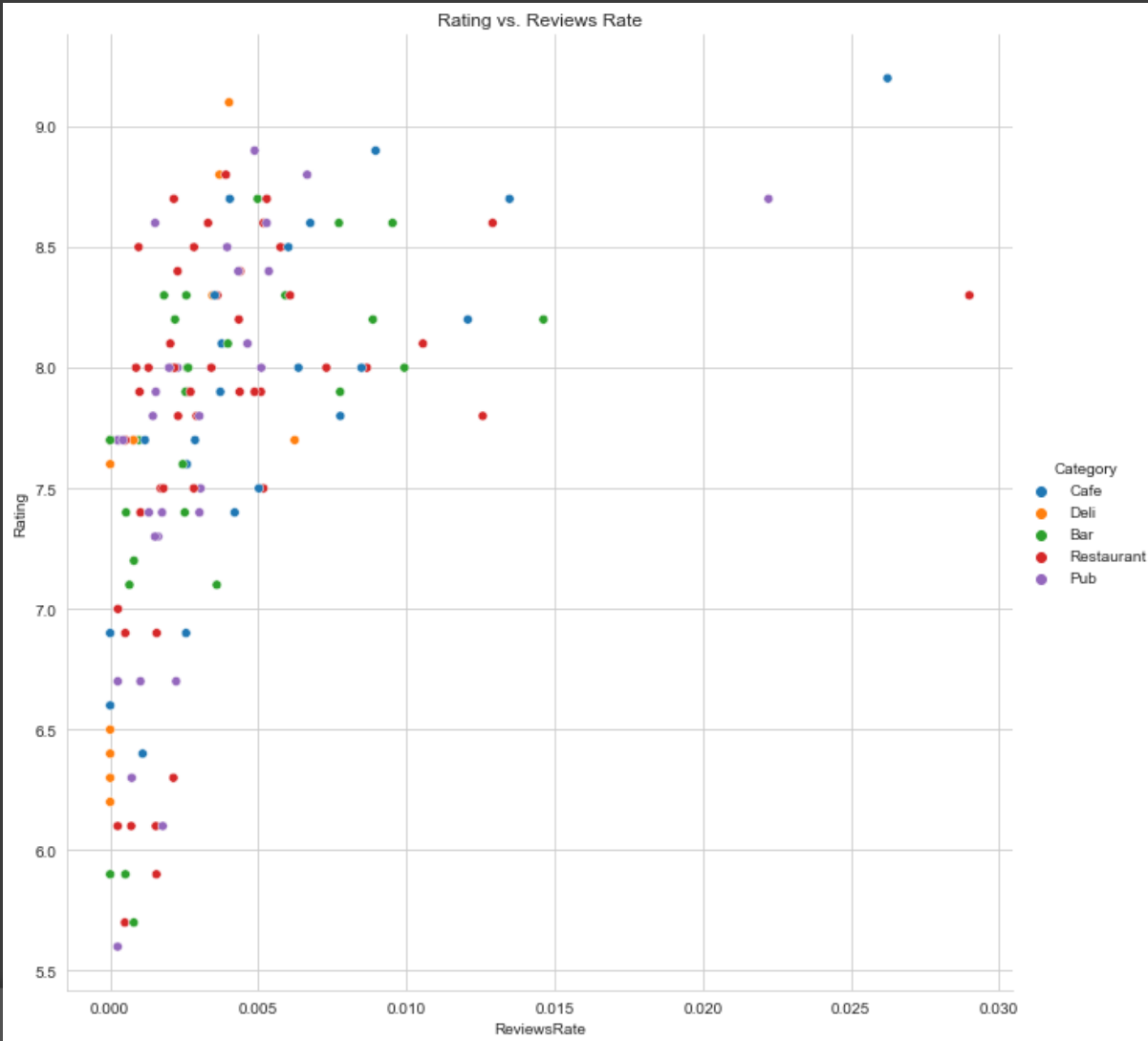
Reviews vs. Days of Operation



- No strong relationship present

Exploratory Data Analysis

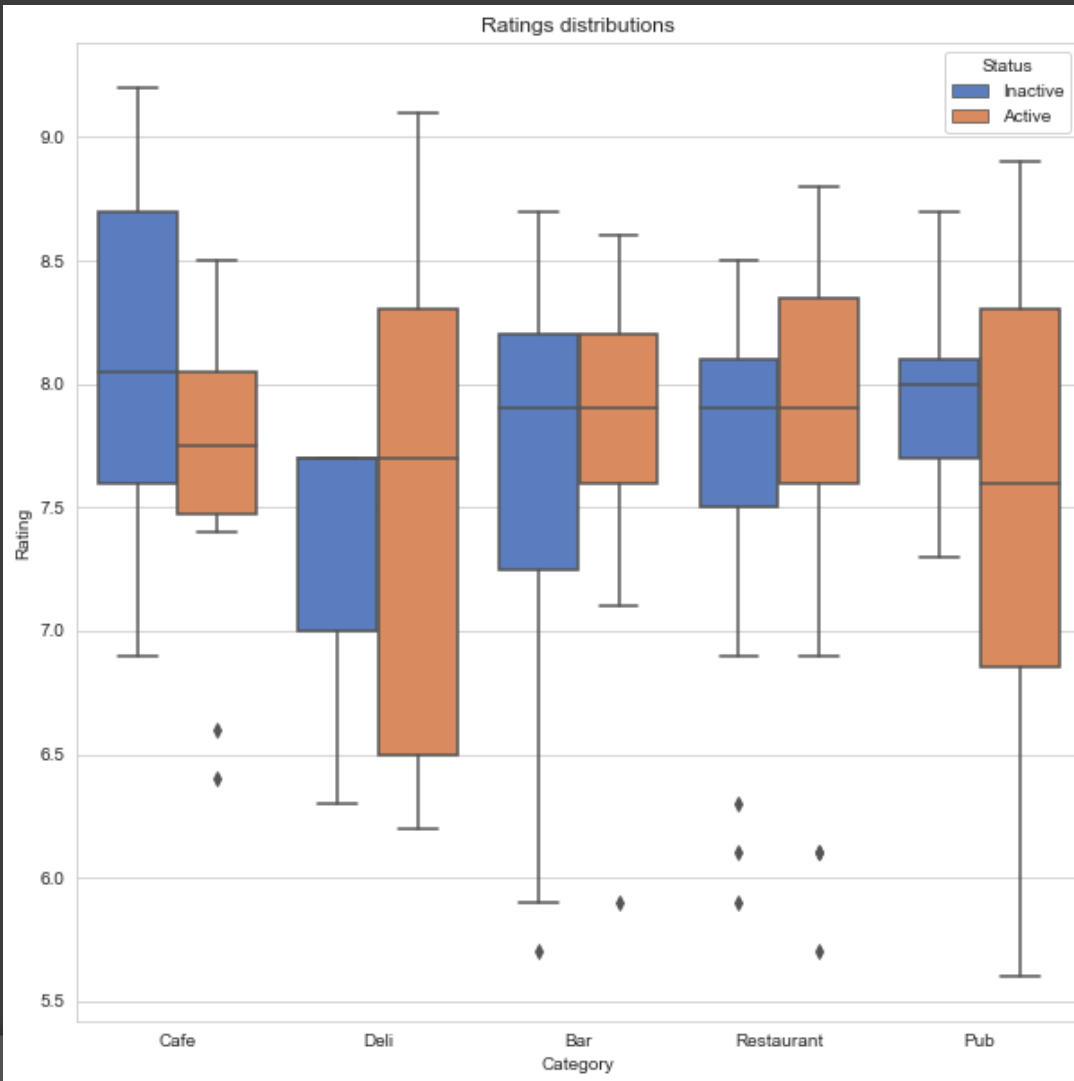
Ratings vs. Reviews (per day)



- Venues which have high ratings also receive more reviews
- This suggests that when customer are particularly satisfied they are also more likely to leave a review

Exploratory Data Analysis

Ratings vs. Status vs. Category



- No apparent trends between the ratings and whether the venues are still in the business (Active) or out of business (Inactive)

Predictive Modelling

- ⦿ Inputs

- Category, Location, Rating, Reviews Rate

- ⦿ Target

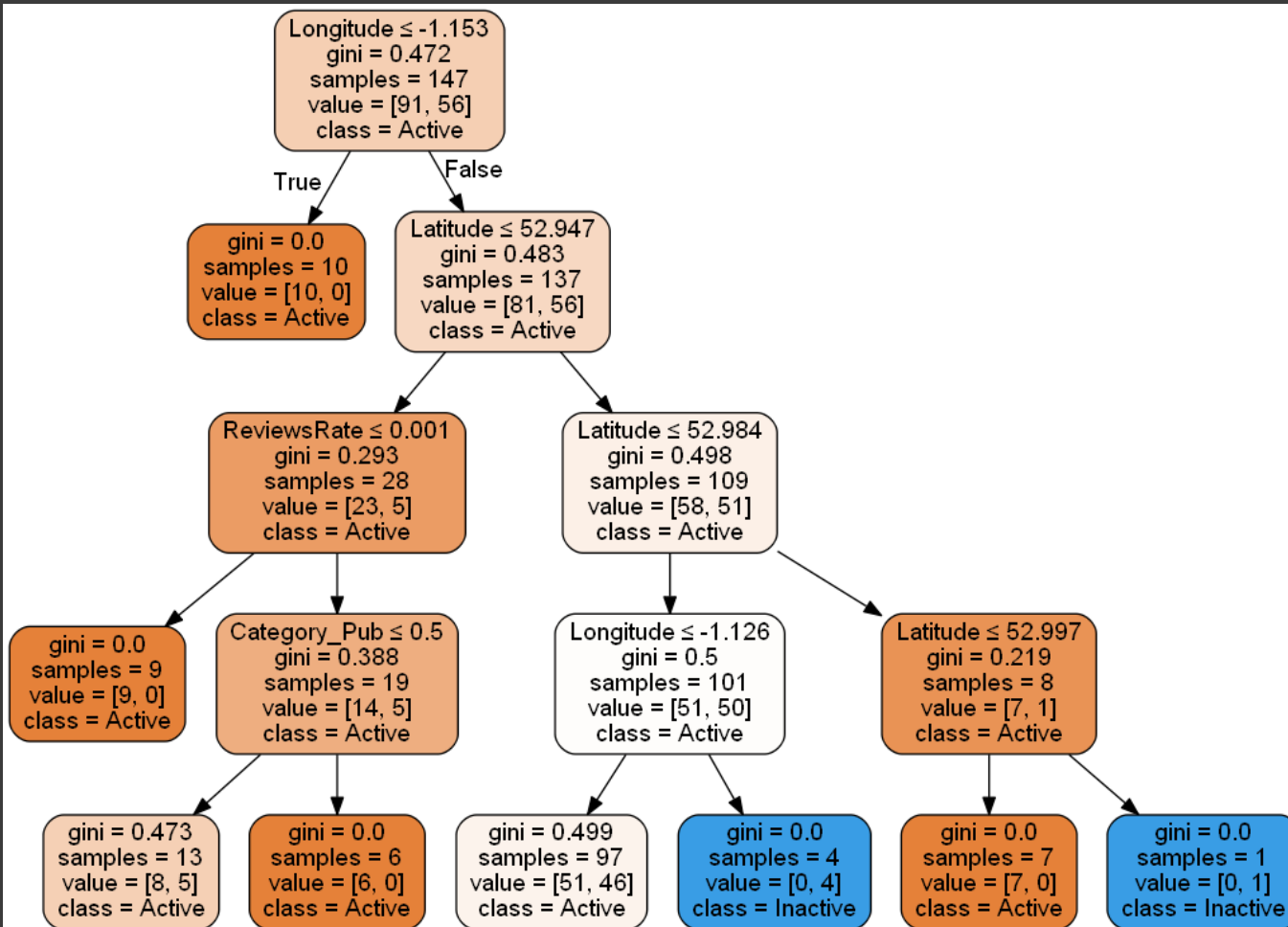
- Status (Active / Inactive)

- ⦿ Classification using decision trees

- Different algorithms with repeated cross-validation
- Overall model accuracy proved poor

Predictive Modelling

Decision Tree Classifier



- Accuracy only in order of 60%

Summary

- Nottingham food venues data extracted from Foursquare API
- Various parameter relationships explored successfully
- Unable to successfully predict which locations and venue types are more likely to stay in business