



Predicting Las Vegas Restaurant Inspections

Lorenz Madarang

Supervised Learning Capstone

July 2018

Research Questions

- ❖ What restaurant variables help predict that a restaurant will receive a downgrade in grade after an inspection?
 - ❖ Past History of Downgrades?
 - ❖ Zip Code of Restaurant?
 - ❖ Category of Restaurant?

Dataset

- ❖ Data retrieved from the Las Vegas Government Open Data website <https://opendata.lasvegasnevada.gov/Public-Safety/Restaurant-Inspections/q8ye-5kww>
- ❖ 159253 restaurant inspections from 2010 to 2018
- ❖ 21297 unique restaurants

Feature Engineering

- Target Value
 - Used RegEx to find 'Downgrade' in 'Inspection Result'
 - Created Boolean 'Downgrade' Column based on RegEx Match method
- Converted 'Inspection Date' and 'Inspection Time' to DateTime Object
- Cumulative Downgrade Count
 - Sorted Data so Oldest Inspection Date was at the top
 - Used lambda function to create cumulative count for each restaurant

Feature Engineering

Inspection Type	Inspection Demerits	Inspection Grade	Permit Status	Inspection Result	Violations	Record Updated	Location 1	Downgrade	Cum_Cou
Routine Inspection	5	A	NaN	Compliant	31,32,35,36,37	02/21/2013 10:26:12 PM	(36.23975, 115.0566759)	0	0.0
Routine Inspection	28	C	NaN	'C' Downgrade	209,211,214,216,217,218,223,225,227,230,232,233	02/21/2013 10:26:12 PM	(36.23975, 115.0566759)	1	1.0
Routine Inspection	0	A	NaN	Compliant	NaN	02/21/2013 10:26:12 PM	(36.23975, 115.0566759)	0	1.0
Routine Inspection	31	C	NaN	'C' Downgrade	208,209,213,214,217,218,220,221,228,230,233	02/21/2013 10:26:12 PM	(36.23975, 115.0566759)	1	2.0
Routine Inspection	0	A	NaN	Compliant	NaN	02/21/2013 10:26:12 PM	(36.23975, 115.0566759)	0	2.0
Routine Inspection	5	A	NaN	Compliant	202	04/25/2013 03:44:31 PM	(36.23975, 115.0566759)	0	2.0
Routine Inspection	0	A	NaN	Compliant	2928,2930	11/26/2013 09:41:14 AM	(36.23975, 115.0566759)	0	2.0
Routine Inspection	9	A	NaN	Compliant	215,216,2928,2930,2955	06/13/2014 12:41:51 PM	(36.23975, 115.0566759)	0	2.0
Routine Inspection	17	B	NaN	'B' Downgrade	207,215,216,218,2912,2928,2930,2955	04/14/2015 09:32:37 AM	(36.23975, 115.0566759)	1	3.0
Re-inspection	3	A	NaN	Compliant	2909,2925	04/28/2015 03:26:00 PM	(36.23975, 115.0566759)	0	3.0
Routine Inspection	0	B	NaN	'B' Downgrade	NaN	03/14/2016 03:54:02 PM	(36.23975, 115.0566759)	1	4.0

Exploratory Data Analysis

Percentage of Downgrades

```
In [12]: #Percentage of downgrades
downgrade_percent = (len(downgrade)/len(restaurant)) * 100
print("{:.2F}% of the dataset contains downgrades".format(downgrade_percent))

11.12% of the dataset contains downgrades
```

t-test of Restaurant Population and Downgrade Restaurants

```
In [14]: print('The Mean Current Demerit for the Las Vegas Restaurant population is {:.2F}'.format(restaurant['Current Demerits'].mean()))
print('The Mean Current Demerit for the Non-downgraded restaurants is {:.2F}'.format(no_downgrade['Current Demerits'].mean()))
print('The Mean Current Demerit for the Downgraded restaurants is {:.2F}'.format(downgrade['Current Demerits'].mean()))

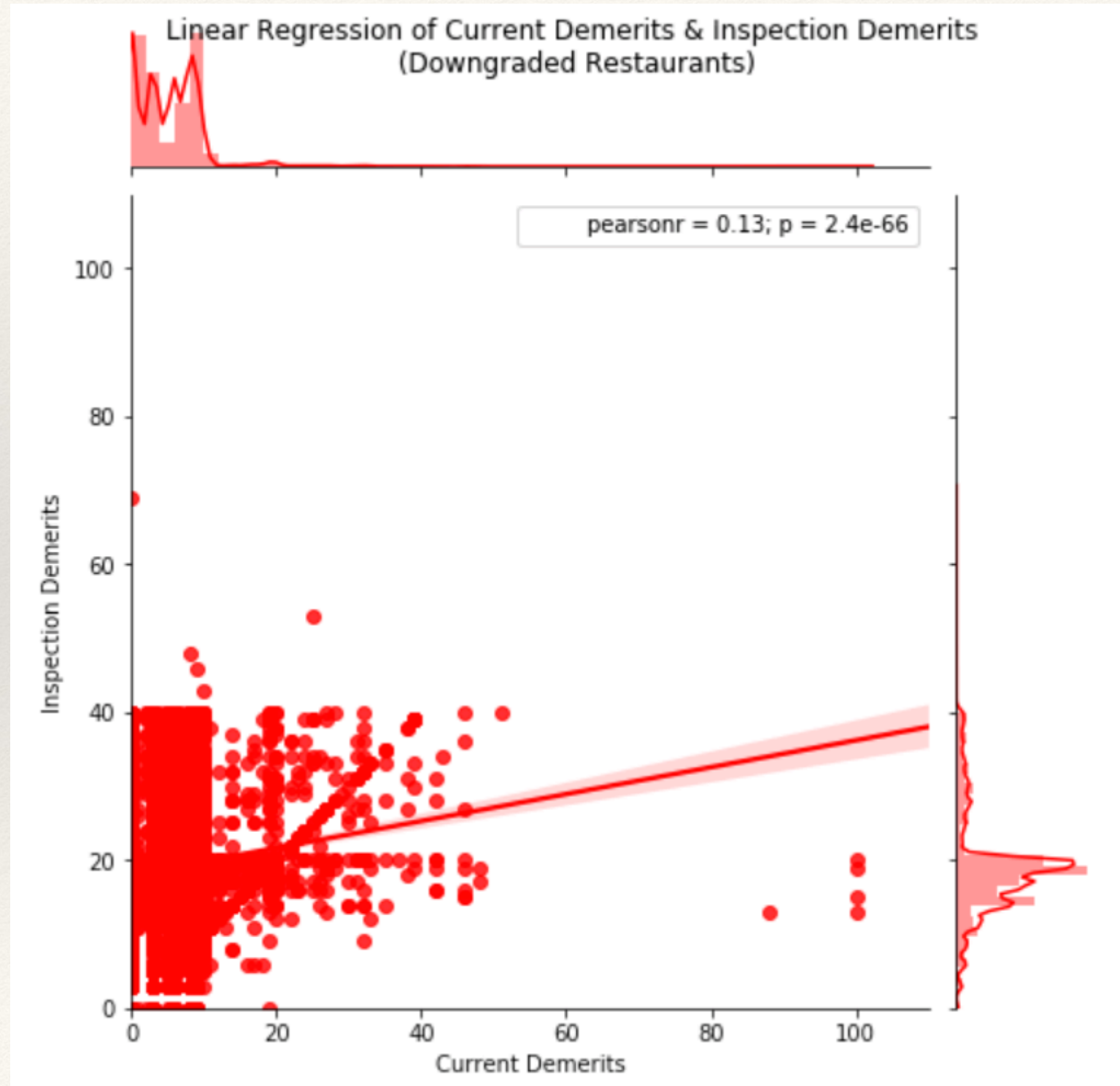
The Mean Current Demerit for the Las Vegas Restaurant population is 3.92
The Mean Current Demerit for the Non-downgraded restaurants is 3.79
The Mean Current Demerit for the Downgraded restaurants is 5.00

In [15]: # t-test between Las Vegas restaurant population and downgraded restaurant sample
from scipy import stats

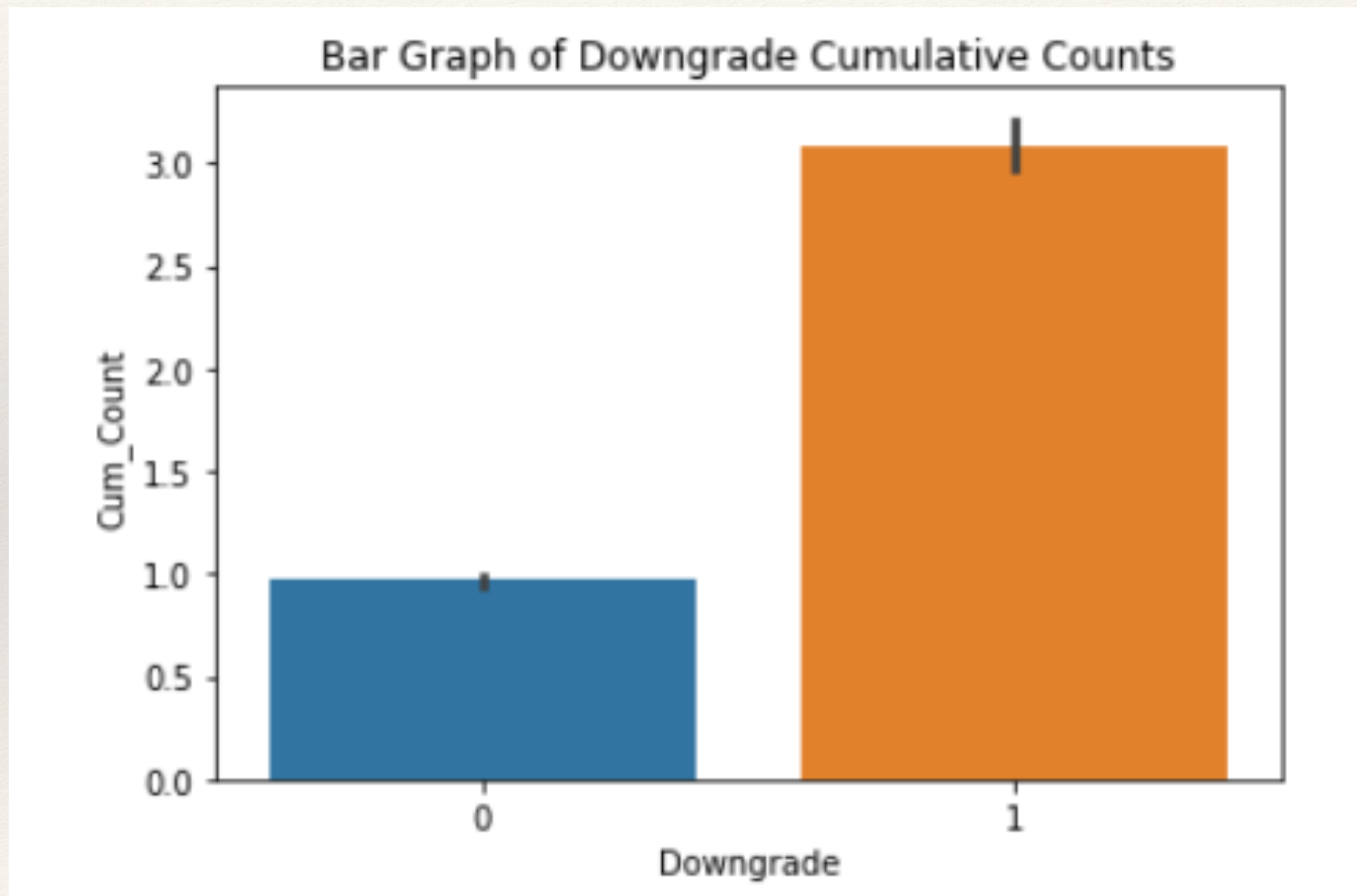
t2, p2 = stats.ttest_ind(x,z)
print("t-value = " + str(t2))
print("p-value = " + str(2*p2))

t-value = -31.4303057648
p-value = 6.17985462055e-216
```


EDA - Current Demerits

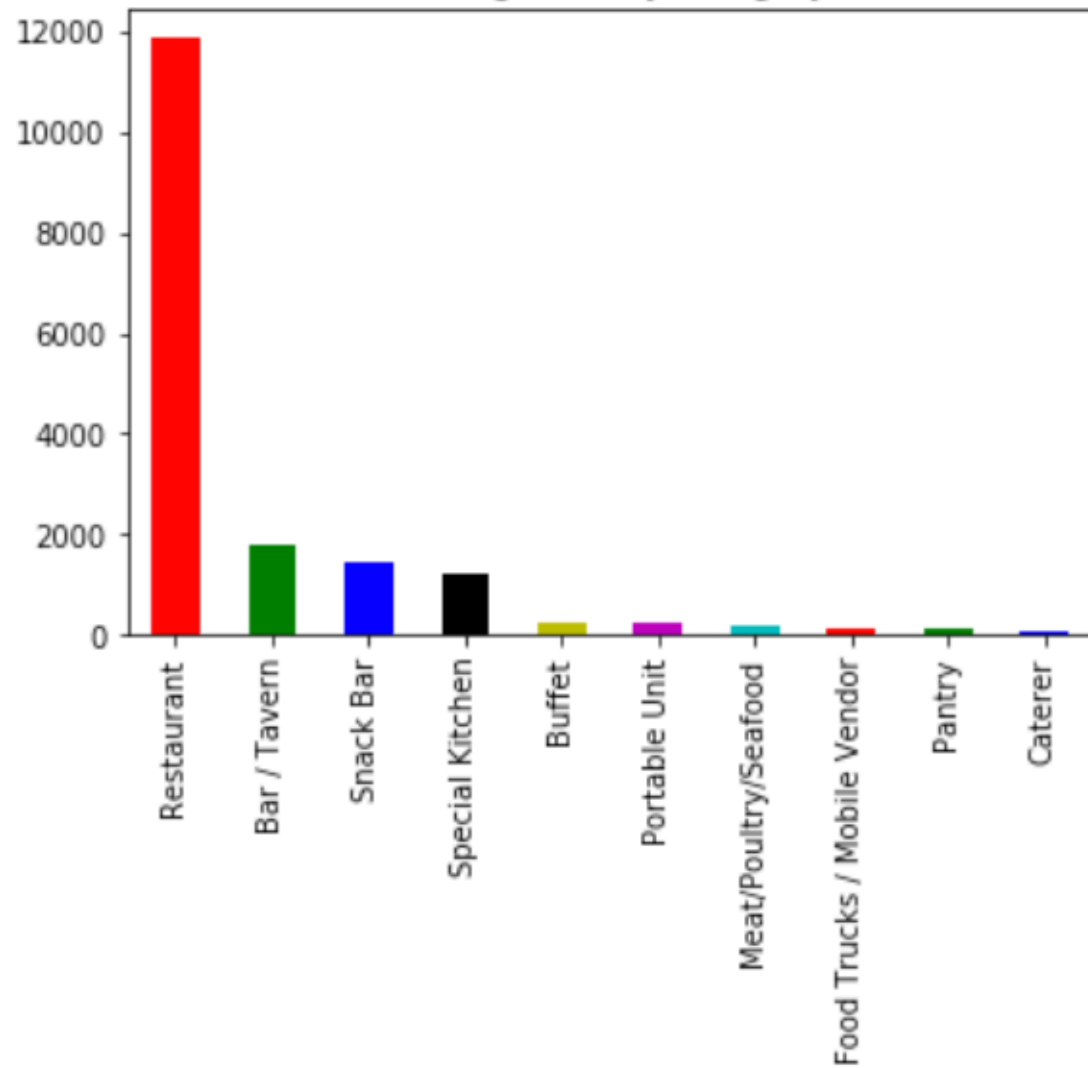


EDA - Downgrade Cumulative Count

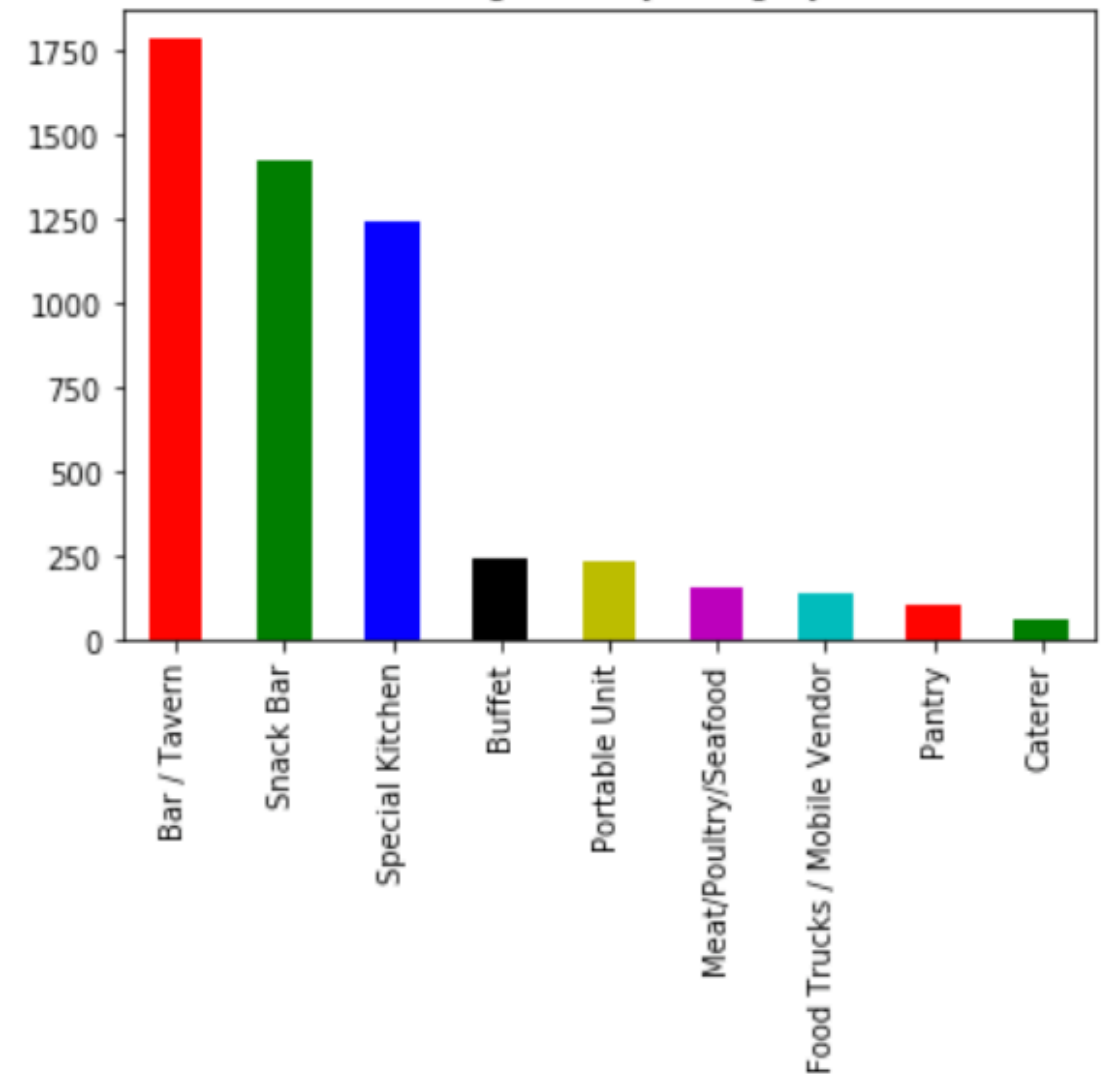


EDA - Count by Category

Downgrades by category

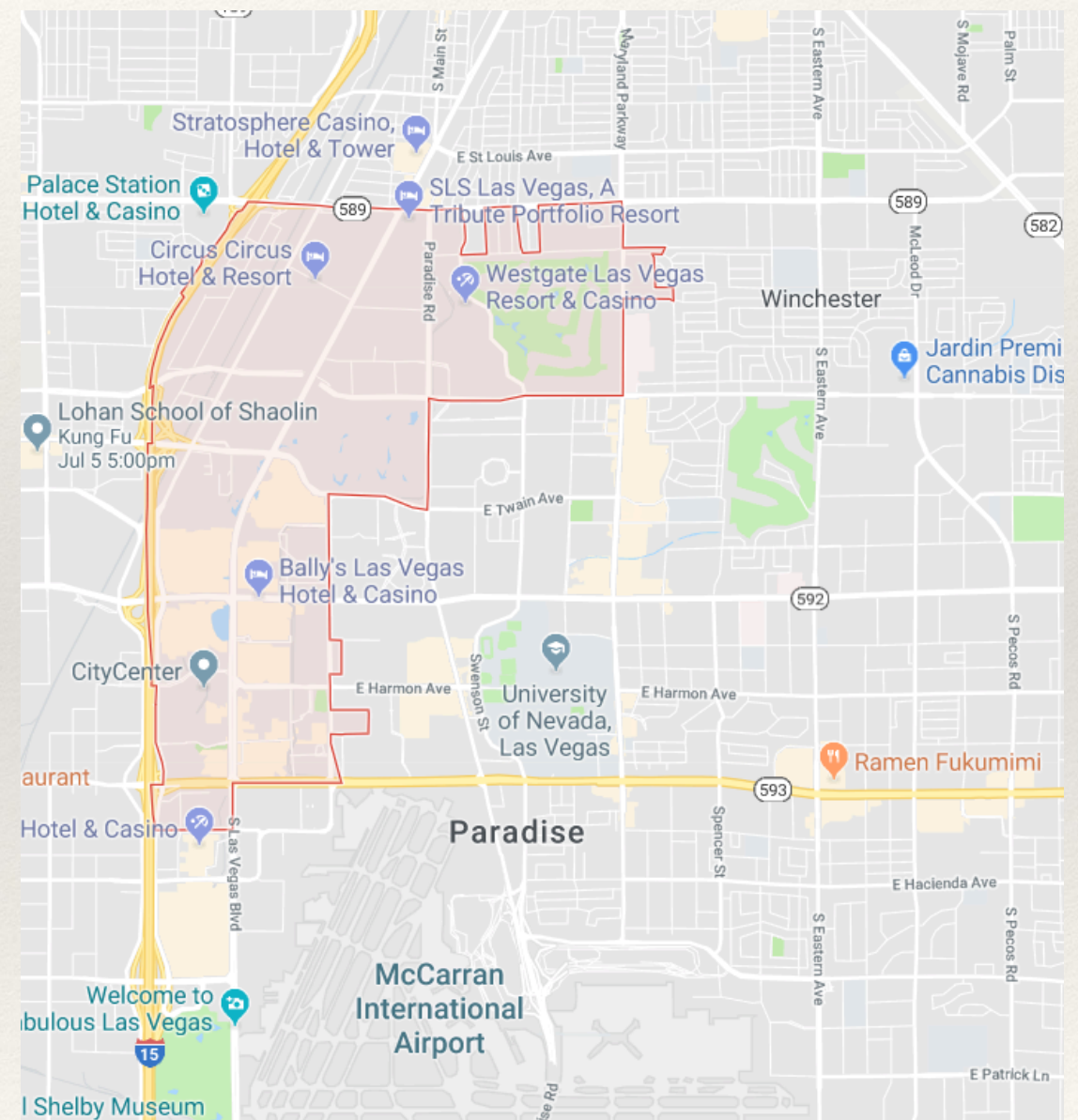
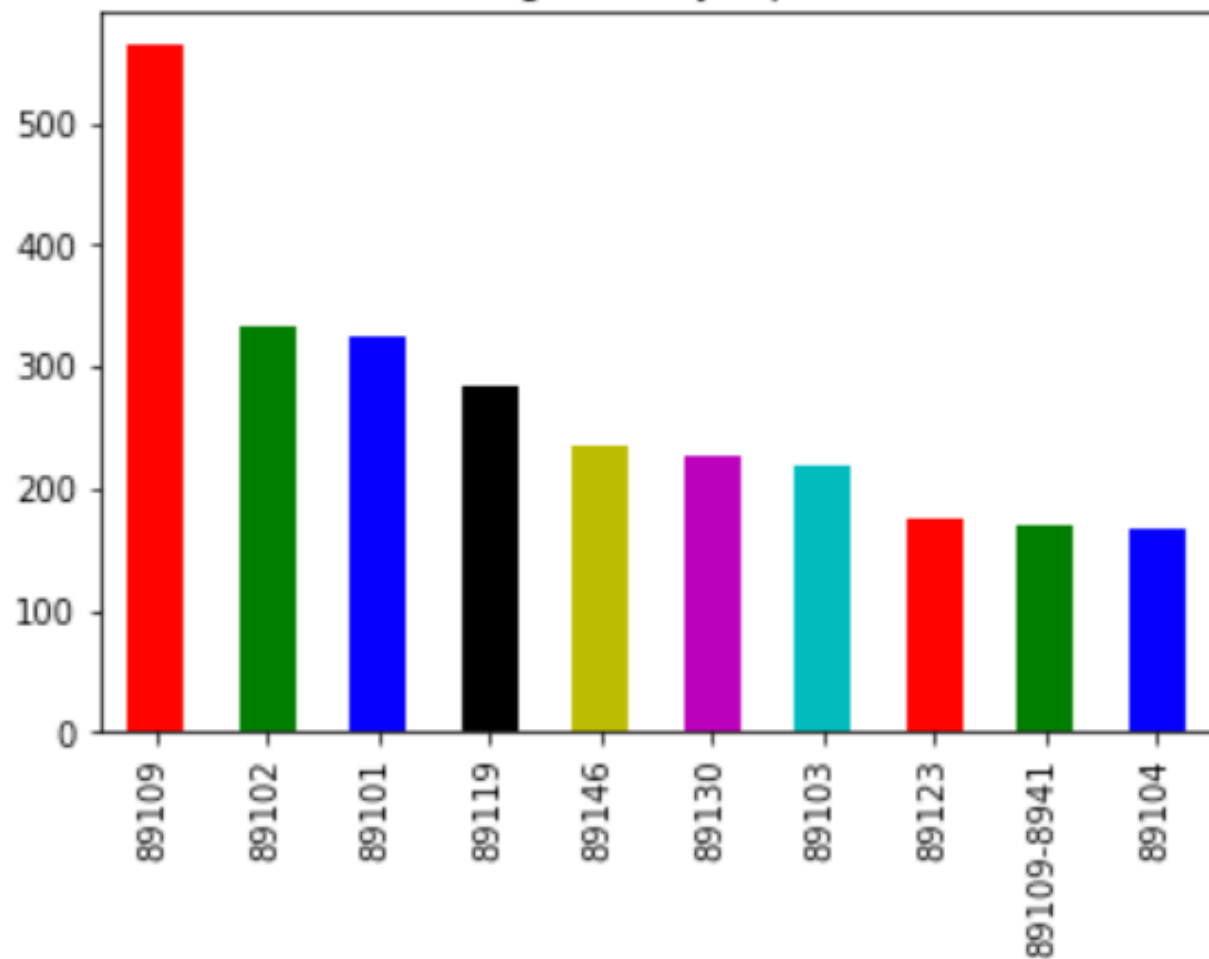


Downgrades by category



EDA - Count by Zip Code

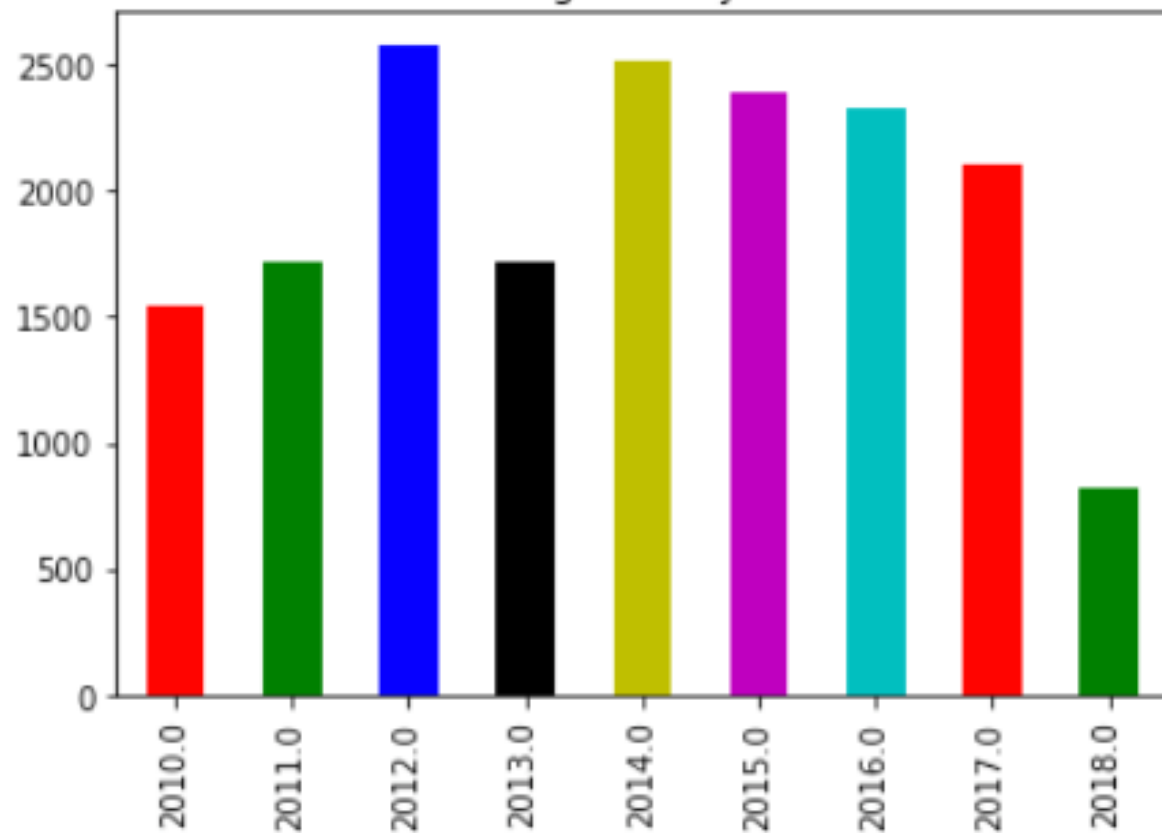
Downgrades by Zip Code



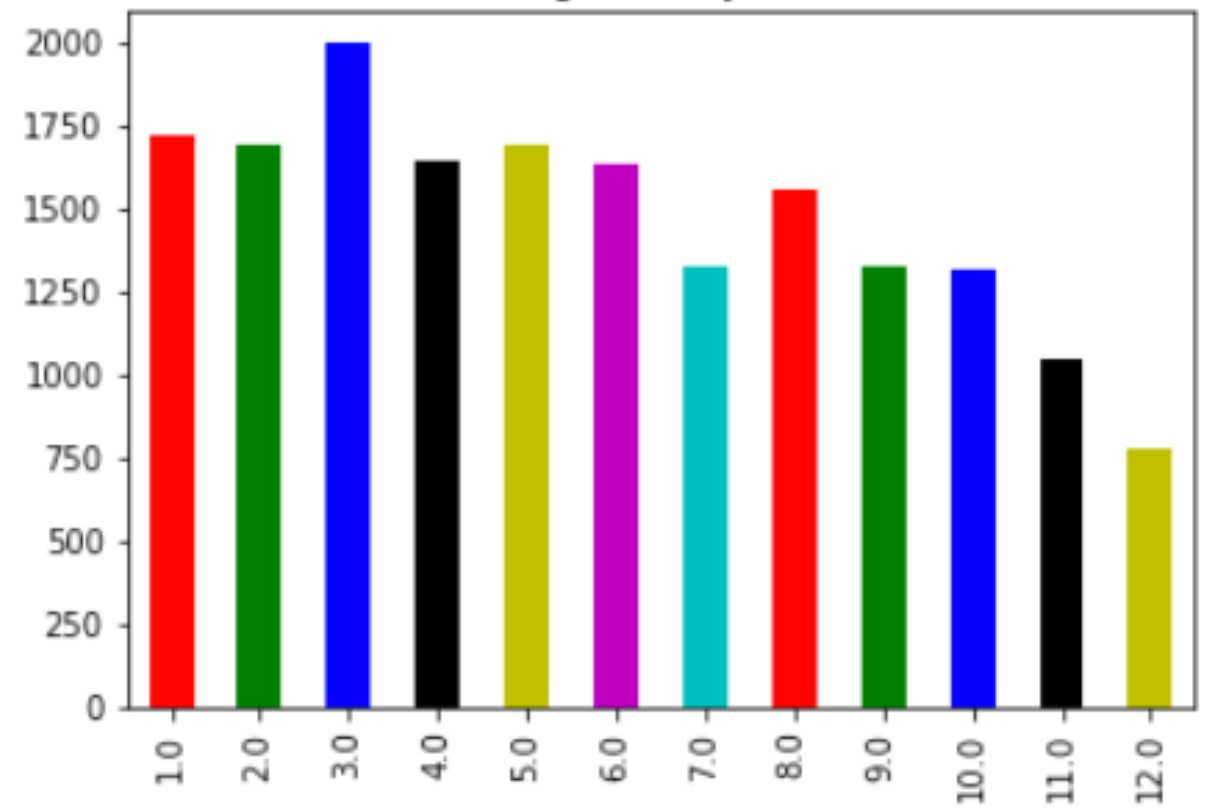


EDA - Count by Year and Month

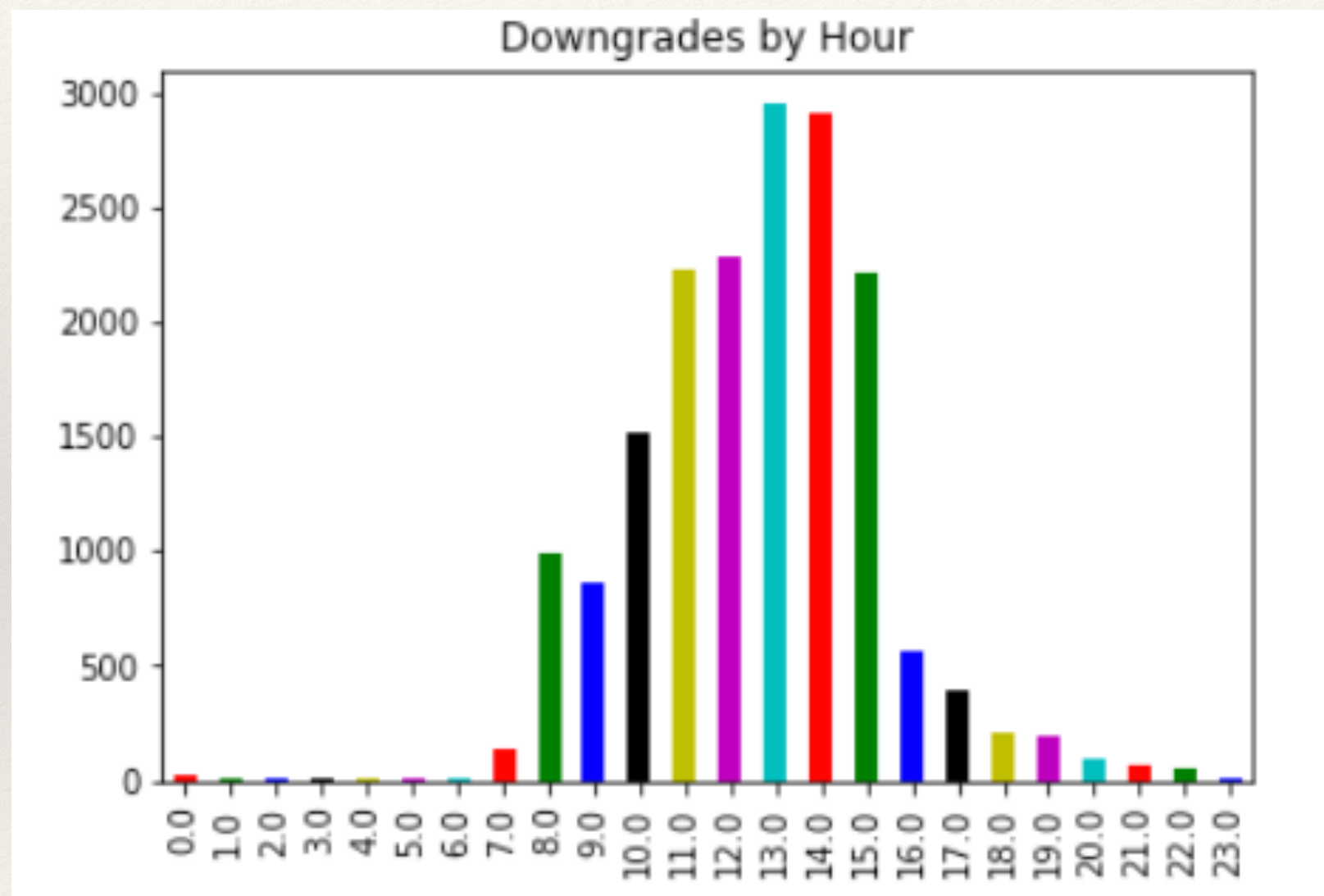
Downgrades by Year



Downgrades by Month



EDA - Count by Hour

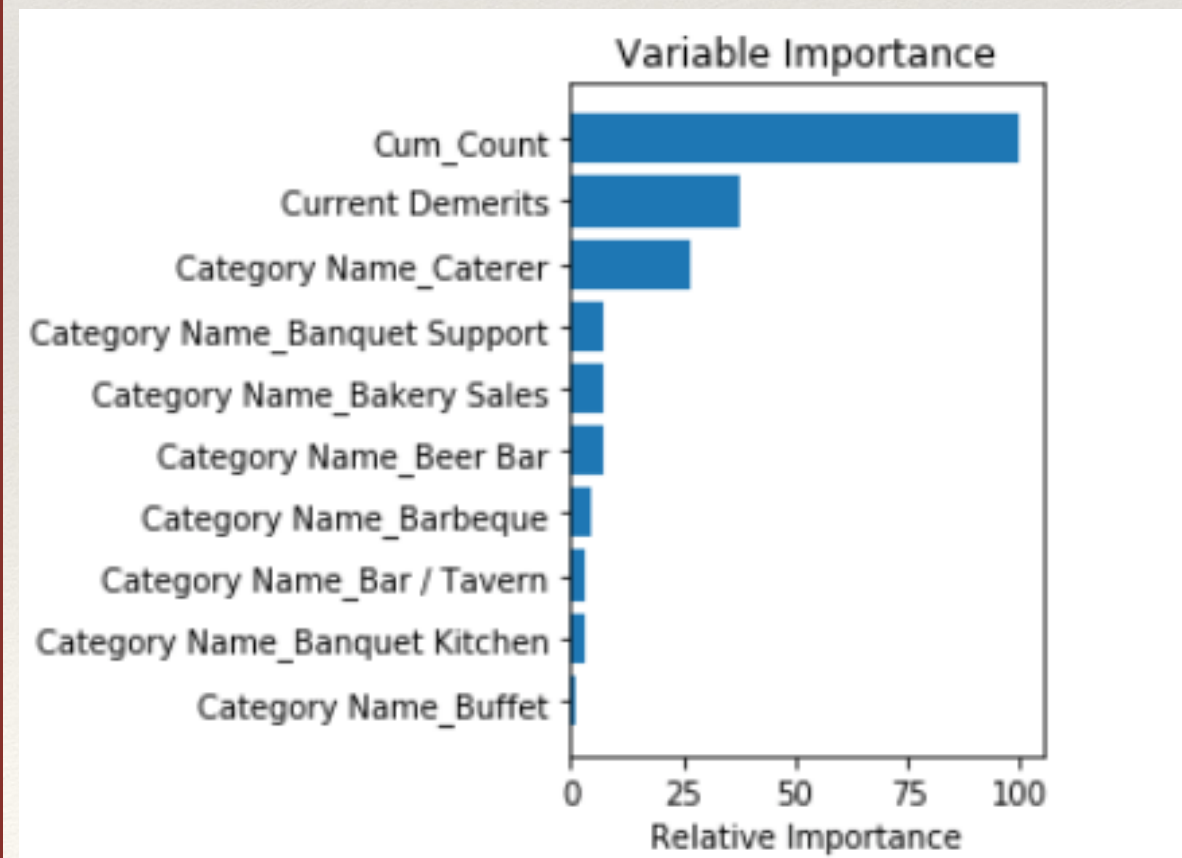


Modeling

- Gradient Boosting Classifier
- Random Forest Classifier
- Super Vector Machine Classifier

Gradient Boosting

- **Default Settings:**
 - Sensitivity: 57.33%
 - Specificity: 89.29%
- **Grid Search Optimized Settings:**
 - Sensitivity: 60.28%(+/-23)
 - Specificity: 89.03%(+/-0.0)



Random Forest

- **Default Settings:**
 - Sensitivity: 51.36%
 - Specificity: 89.50%
- **Grid Search Optimized Settings:**
 - Sensitivity: 65.9%(+/-23)
 - Specificity: 88.9%(+/-0.0)

Support Vector Machine

- **Default Settings:**
 - Sensitivity: 53.60%
 - Specificity: 89.17%
- **Optimized Setting:**
 - Sensitivity: 58.7%(+/-23)
 - Specificity: 88.8%(+/-0.0)

Model Analysis Summary

	Model	Mean_Accuracy_Training	Mean_Accuracy_Test	Mean_Precision_Train	Mean_Precision_Test
1	Random_Forest	0.891	0.889	0.636	0.659
0	Gradient_Boost	0.635	0.603	0.635	0.603
2	Support_Vector	0.891	0.888	0.571	0.587

Conclusion

- ❖ Best Model is Random Forest Classifier
- ❖ Important Variables
 - ❖ History of Downgrades
 - ❖ Current Demerits
 - ❖ Category of Restaurant
- ❖ Machine-Learning Optimized Restaurant selection can find more downgrades