

Boulder County Health Inspection Scores



vs



Google Business Ratings

Is there a correlation between restaurant health inspection scores and Google rating?

Are there particular violations that have a bigger impact on Google rating than others?



Credit: <https://chicago.eater.com/2021/7/29/22600181/tamale-guy-claudio-velez-chicago-bar-food-vendor>

"Santa Claus for the drunk and hungry"
https://en.wikipedia.org/wiki/Tamale_Guy

Is there a correlation between restaurant health inspection scores and Google rating?

Are there particular violations that have a bigger impact on Google rating than others?

Inspection Score Scale

0 - 19	Excellent
20 - 39	Good
40 - 69	Fair
70 - 99	Marginal
> 100	Unacceptable

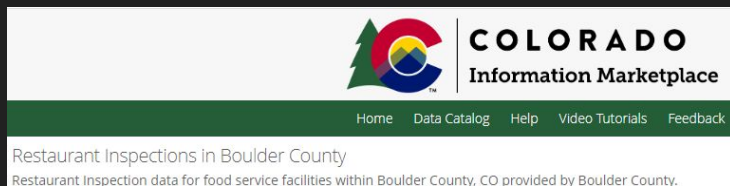
Google Rating Scale





Database

Merging Datasets



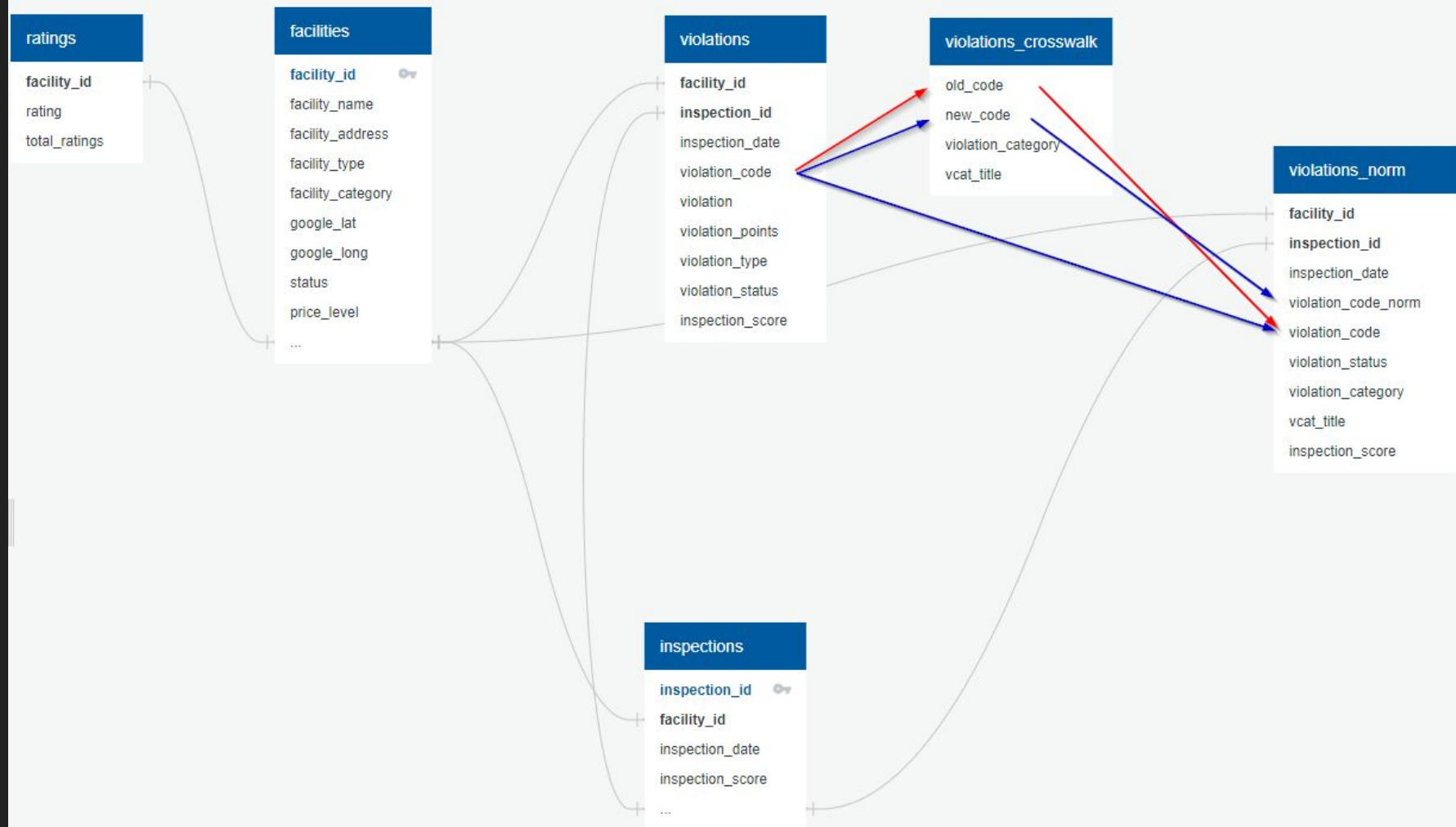
Facility ID
Address
Facility Name
Inspection Date
Inspection Type
Violation



Address
Facility Name
Coordinates

Address
Facility Name
Coordinates
Rating
Total Ratings
Price Level
Status

Facility ID
Address
Facility Name
Inspection Date
Inspection Type
Violation
Coordinates
Rating
Total Ratings
Price Level
Status



Crosswalking Between Old, New Violation Types

violation_category	category_title	new_code	new_code_title	old_code	old_code_title
2	personnel	FC03	Management, food employee and conditional employee; knowledge, responsibilities and reporting	02F	Training Needed
2	personnel	FC04	Proper use of restriction and exclusion	02A	Personnel With Infections Restricted
2	personnel	FC04	Proper use of restriction and exclusion	02B	Wounds Properly Covered
2	personnel	FC04	Proper use of restriction and exclusion	02D	Hygienic Practices
2	personnel	FC06	Proper eating, tasting, drinking, or tobacco use	02E	Smoking Eating Drinking
2	personnel	FC08	Hands clean & properly washed	02C	Hands Washed As Needed
2	personnel	FC09	No bare hand contact with RTE food or a pre-approved alternative procedure properly allowed	02G	No Bare Hand Contact
6	toilets/handwashing	FC10	Adequate handwashing sinks properly supplied and accessible	06B	Accessible
6	toilets/handwashing	FC10	Adequate handwashing sinks properly supplied and accessible	06C	Soap and Drying Devices
1	food source	FC11	Food obtained from approved source	01A	Approved Source
1	food source	FC13	Food in good condition, safe, & unadulterated	01B	Wholesome Free of Spoilage
9	food labeling/protection	FC15	Food separated and protected	09B	Food Protected from Contamination
12	equipment/utensil cleaning	FC16	Food contact surfaces; cleaned & sanitized	04A	Manual
12	equipment/utensil cleaning	FC16	Food contact surfaces; cleaned & sanitized	12A	Food Contact Surfaces Cleaning
3	food temperature control	FC18	Proper cooking time & temperatures	03D	Required Cooking Temperature
3	food temperature control	FC19	Proper reheating procedures for hot holding	03B	Rapidly Reheat to 165 Degrees or Greater

Pivot Tables!

```
139 select * from binary_violations_pivot;
```

```
140
```

```
141
```

Data Output		Explain	Messages	Notifications										
	facility_id character varying	fc01 integer	fc02 integer	fc03 integer	fc04 integer	fc05 integer	fc06 integer	fc08 integer	fc09 integer	fc10 integer	fc11 integer	fc13 integer	fc14 integer	fc15 integer
1	FA0000009	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	1	[null]	[null]	[null]	[null]
2	FA0000010	[null]	[null]	[null]	1	[null]	1	[null]	1	1	[null]	[null]	[null]	1
3	FA0000011	[null]	[null]	[null]	1	[null]	1	[null]	[null]	[null]	[null]	[null]	[null]	[null]
4	FA0000015	[null]	[null]	[null]	1	[null]	1	1	[null]	[null]	[null]	1	[null]	1
5	FA0000017	[null]	[null]	[null]	[null]		[null]	1	[null]	1	[null]	[null]	[null]	[null]
6	FA0000018	[null]	[null]	1	[null]	[null]	[null]	1	[null]	[null]	[null]	[null]	[null]	[null]
7	FA0000021	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
8	FA0000024	[null]	1	[null]	1	1	[null]	[null]	1	1	[null]	[null]	[null]	1
9	FA0000025	[null]	1	[null]	1	1	1	1	1	[null]	[null]	[null]	[null]	[null]
10	FA0000026	[null]	[null]	[null]	[null]	[null]	1	[null]	[null]	1	[null]	[null]	[null]	[null]

```
1 select * from violation_cat_counts_pivot;
```

Data Output		Explain	Messages	Notifications									
	facility_id character varying	cat_1 integer	cat_2 integer	cat_3 integer	cat_5 integer	cat_6 integer	cat_7 integer	cat_8 integer	cat_9 integer	cat_10 integer	cat_12 integer	cat_13 integer	cat_14 integer
1	FA0000009	1	[null]	[null]	1	2	[null]	[null]	[null]	[null]	1	[null]	[null]
2	FA0000010	1	2	1	2	1	[null]	[null]	3	1	[null]	[null]	1
3	FA0000011	[null]	2	2	1	[null]	[null]	1	[null]	[null]	[null]	[null]	[null]
4	FA0000015	1	2	2	1	[null]	[null]	[null]	1	1	1	2	[null]
5	FA0000017	1	1	[null]	2	1	[null]	[null]	[null]	[null]	1	[null]	[null]
6	FA0000018	3	1	2	1	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
7	FA0000021	[null]	[null]	[null]	[null]	[null]	[null]	1	[null]	[null]	[null]	[null]	[null]
8	FA0000024	1	3	1	1	2	1	[null]	2	[null]	1	[null]	[null]
9	FA0000025	[null]	4	2	1	[null]	[null]	1	[null]	[null]	1	[null]	[null]
10	FA0000026	[null]	1	1	1	1	[null]	2	1	[null]	1	1	[null]

Machine Learning

```
In [30]: 1 # Define the logistic regression model
2 log_classifier = LogisticRegression(solver="lbfgs",max_iter=200)
3
4 # Train the model
5 log_classifier.fit(X_train,y_train)
6
7 # Evaluate the model
8 y_pred = log_classifier.predict(X_test)
9 print(f" Logistic regression model accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

Logistic regression model accuracy: 0.598

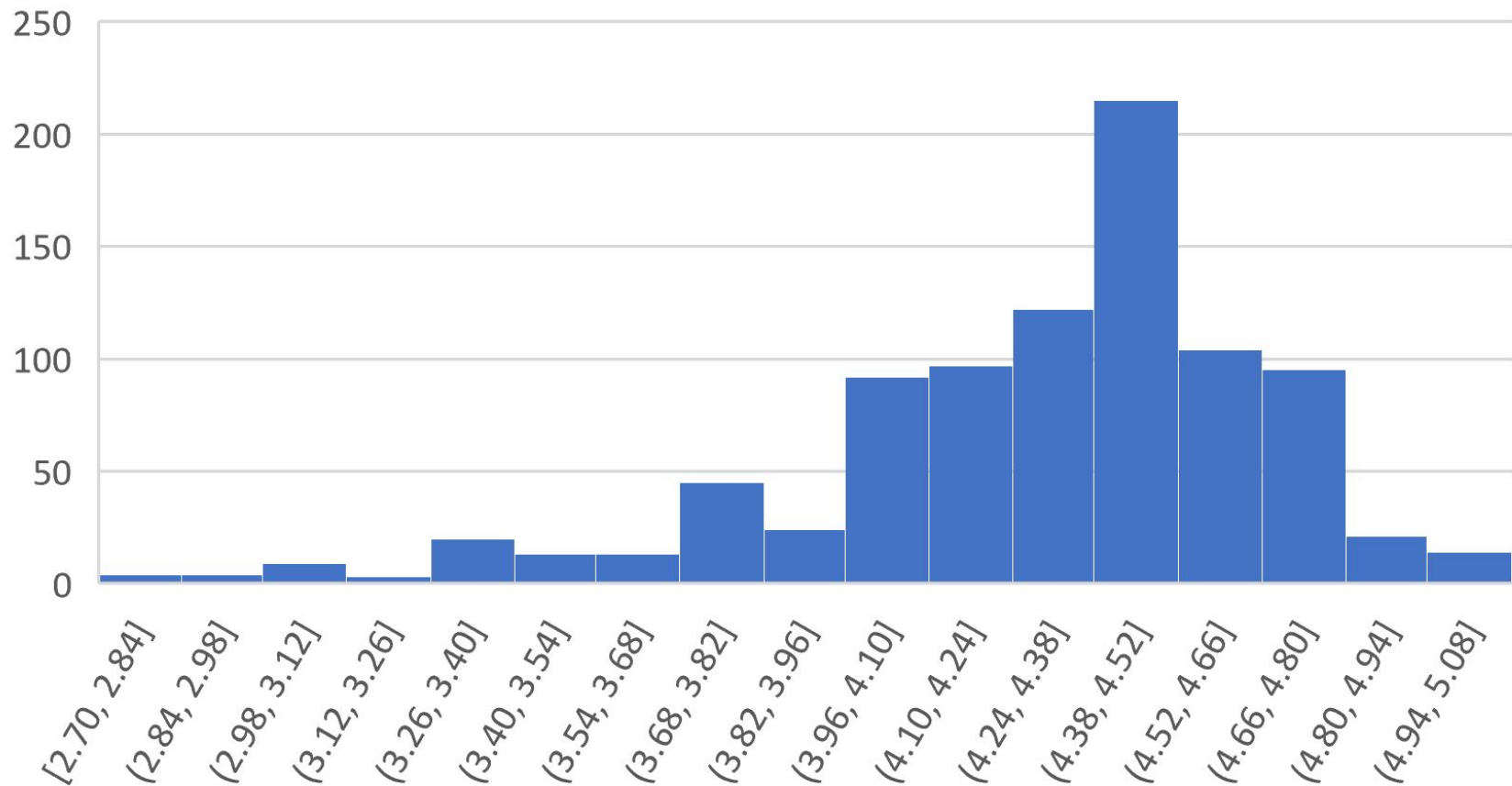
```
In [31]: 1 # Create the SVM model
2 svm = SVC(kernel='linear')
3
4 # Train the model
5 svm.fit(X_train, y_train)
6
7 # Evaluate the model
8 y_pred = svm.predict(X_test_scaled)
9 print(f" SVM model accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

SVM model accuracy: 0.598

```
In [32]: 1 # Create a random forest classifier.
2 rf_model = RandomForestClassifier(n_estimators=128, random_state=78)
3
4 # Fitting the model
5 rf_model = rf_model.fit(X_train_scaled, y_train)
6
7 # Evaluate the model
8 y_pred = rf_model.predict(X_test_scaled)
9 print(f" Random forest predictive accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

Random forest predictive accuracy: 0.558

Distribution of Google Ratings



```
In [34]: 1 # Define the logistic regression model
2 log_classifier = LogisticRegression(solver="lbfgs",max_iter=200)
3
4 # Train the model
5 log_classifier.fit(X_train,y_train)
6
7 # Evaluate the model
8 y_pred = log_classifier.predict(X_test)
9 print(f" Logistic regression model accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

Logistic regression model accuracy: 0.737

```
In [35]: 1 # Create the SVM model
2 svm = SVC(kernel='linear')
3
4 # Train the model
5 svm.fit(X_train, y_train)
6
7 # Evaluate the model
8 y_pred = svm.predict(X_test_scaled)
9 print(f" SVM model accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

SVM model accuracy: 0.737

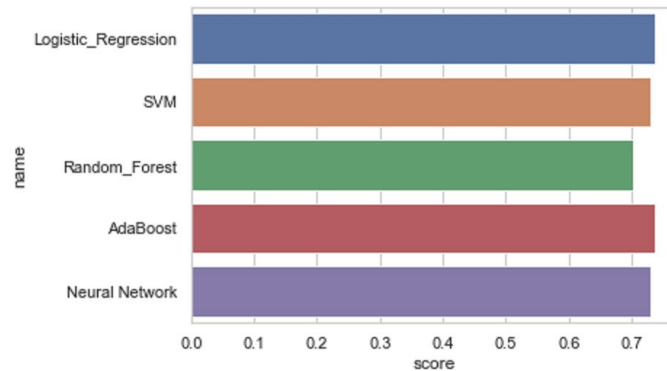
```
In [36]: 1 # Create a random forest classifier.
2 rf_model = RandomForestClassifier(n_estimators=128, random_state=78)
3
4 # Fitting the model
5 rf_model = rf_model.fit(X_train_scaled, y_train)
6
7 # Evaluate the model
8 y_pred = rf_model.predict(X_test_scaled)
9 print(f" Random forest predictive accuracy: {accuracy_score(y_test,y_pred):.3f}")
```

Random forest predictive accuracy: 0.728

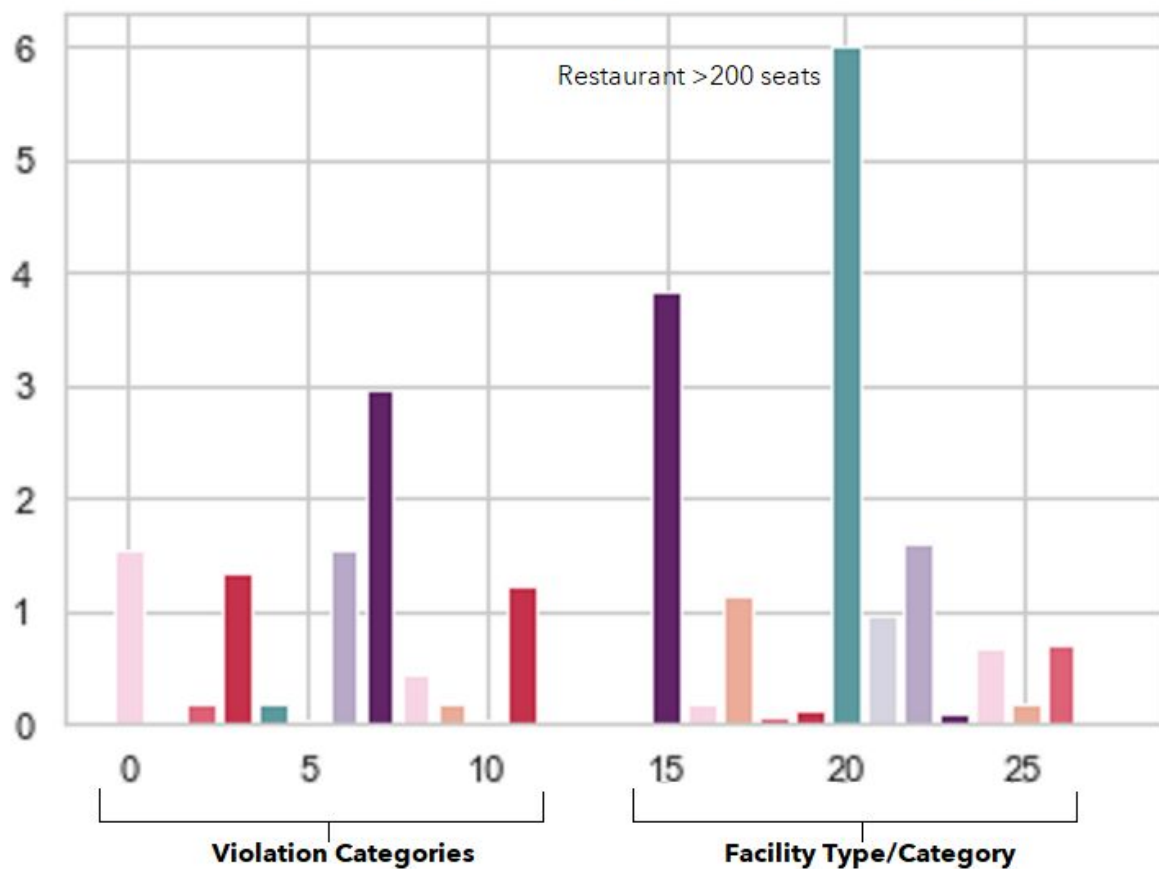
ML Model Comparisons

	name	score
0	Logistic_Regression	0.735632
1	SVM	0.729885
2	Random_Forest	0.701149
3	AdaBoost	0.735632
4	Neural Network	0.729885

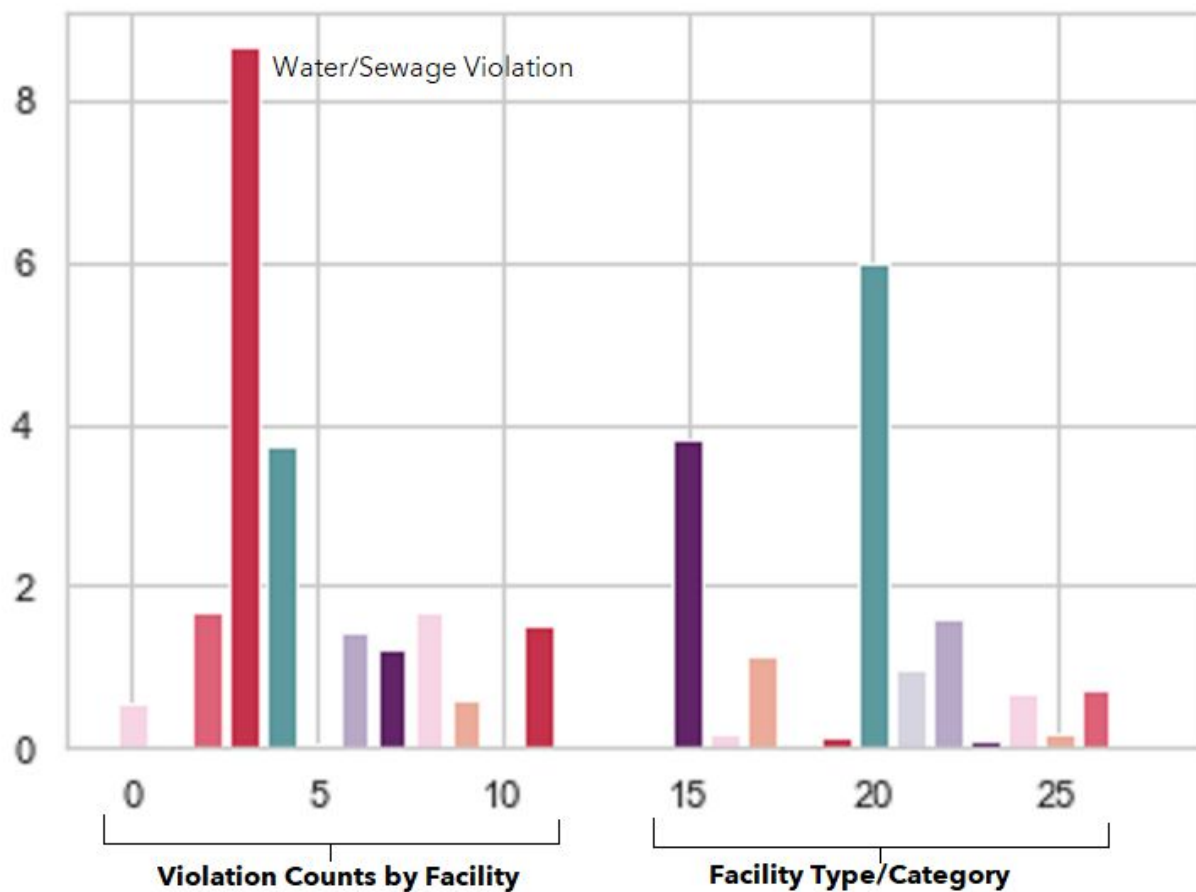
ML Model Comparisons



Feature Selection by Violation Category



Feature Selection by Violation Count



Insert Website Link Here