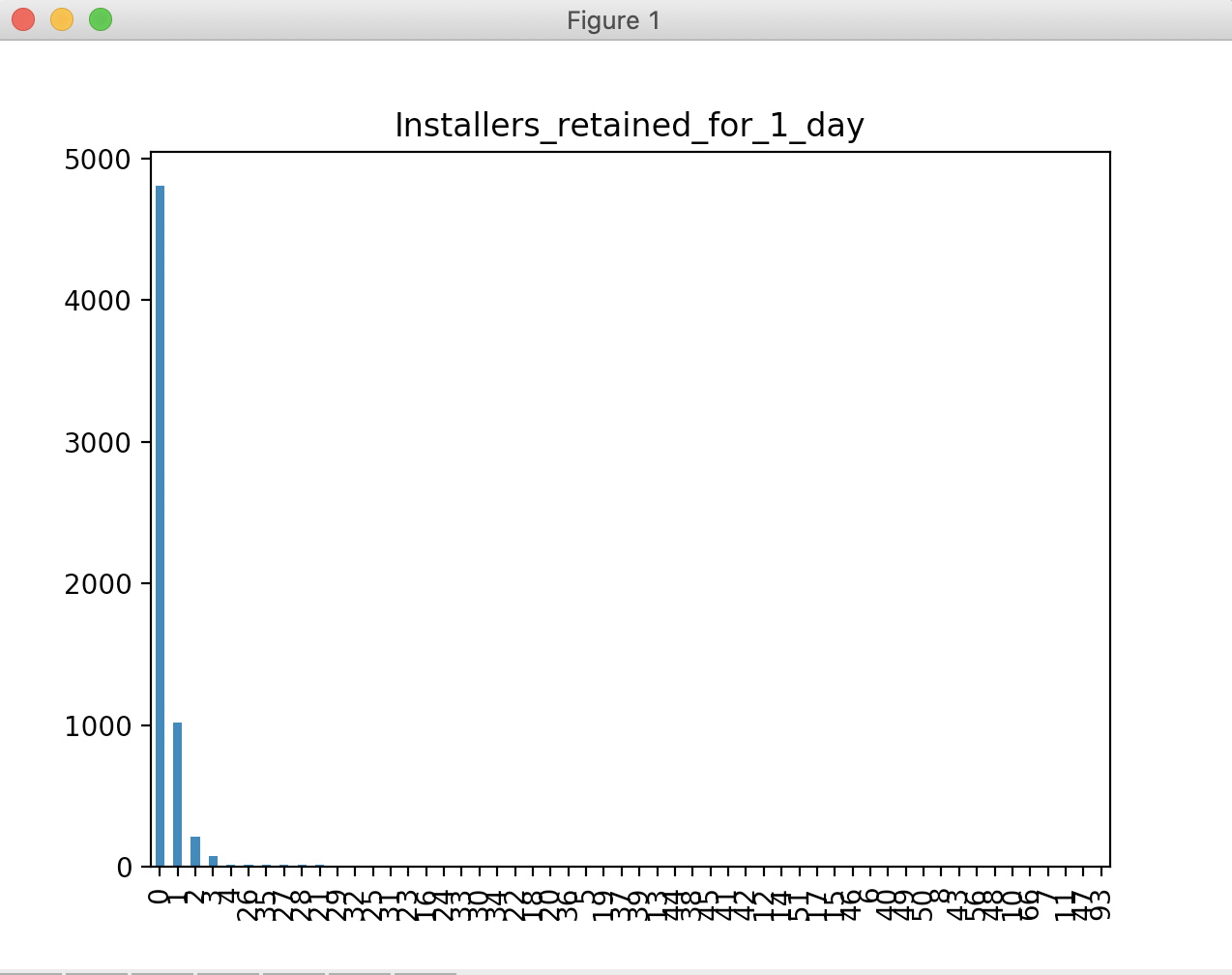
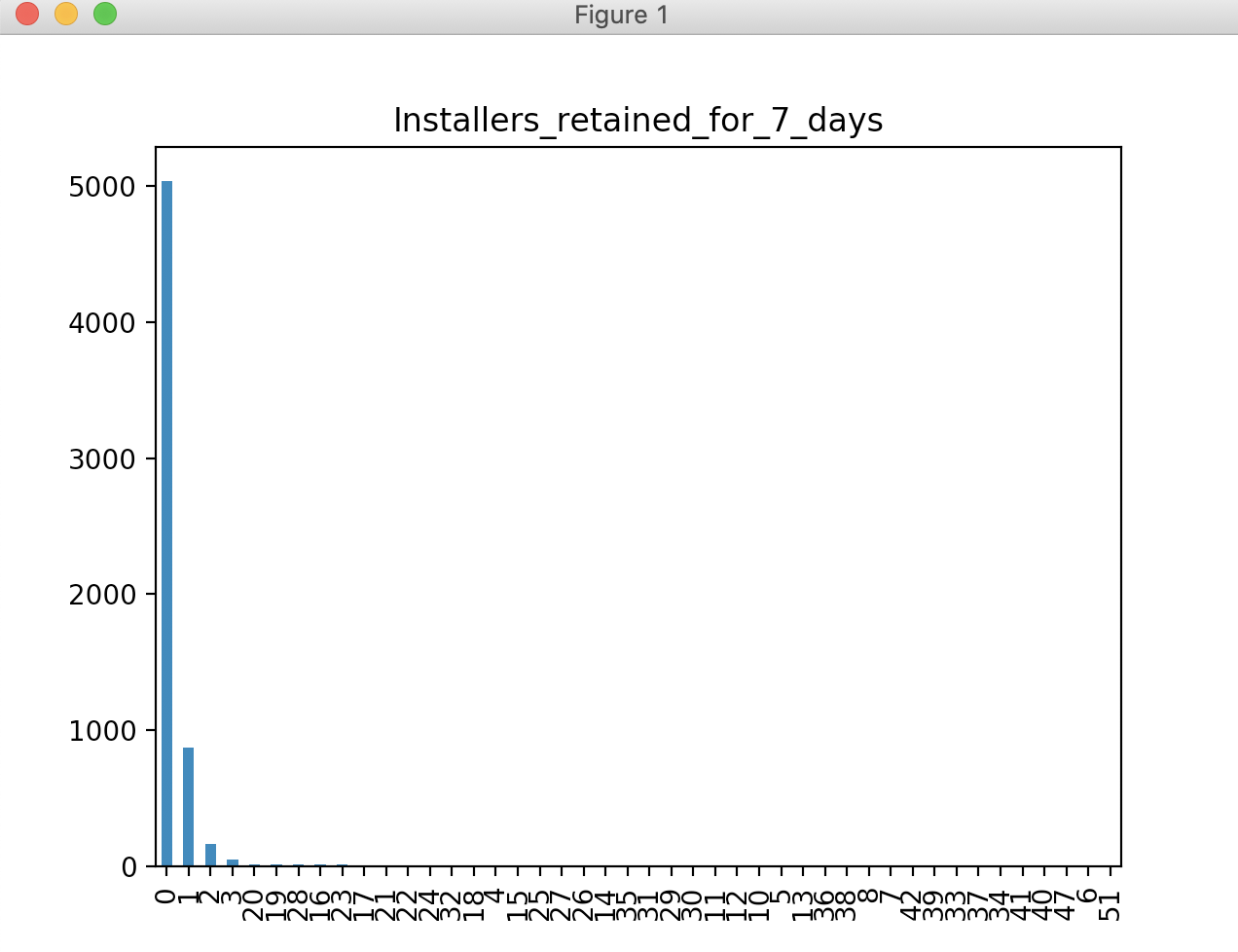
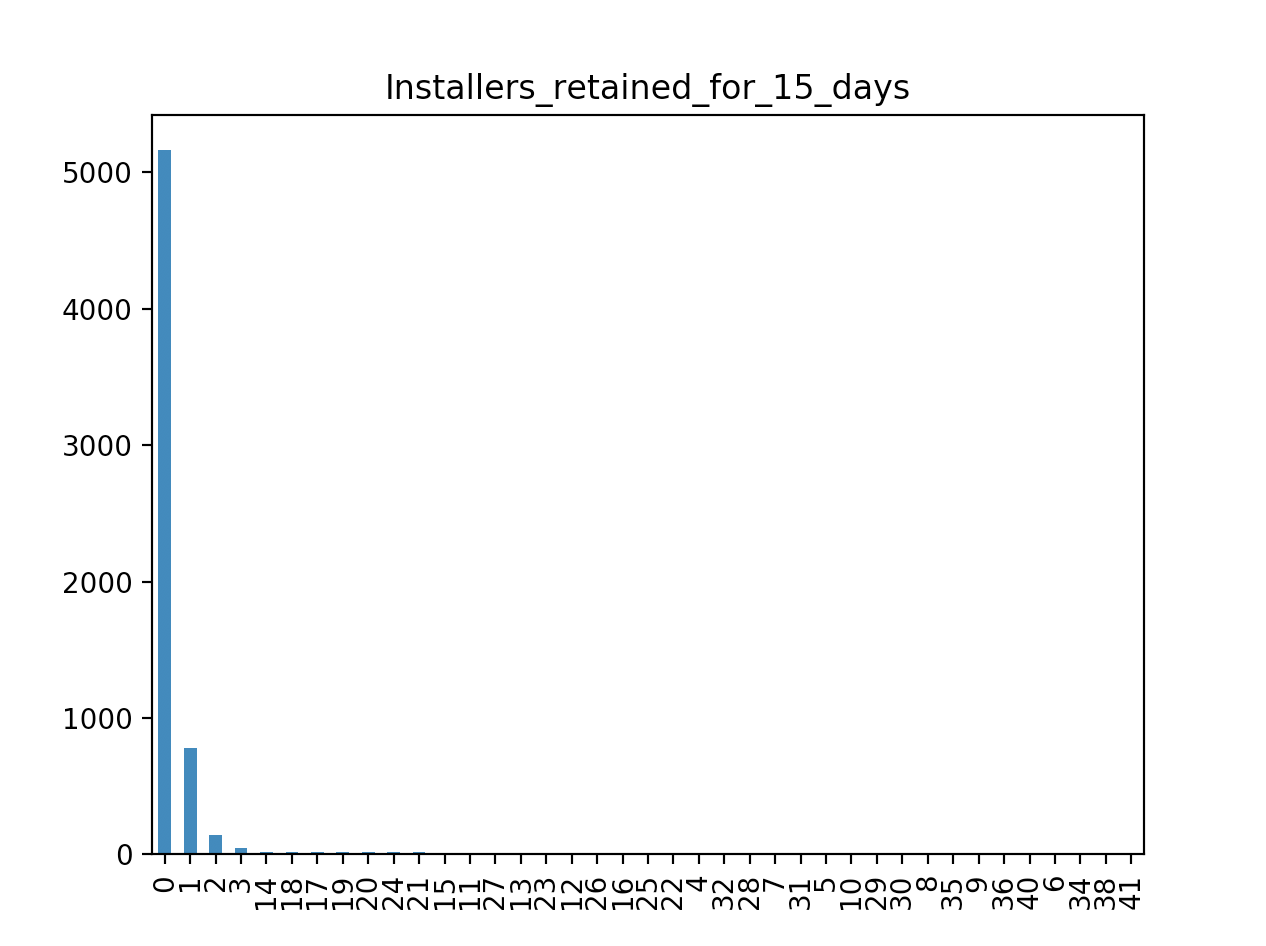
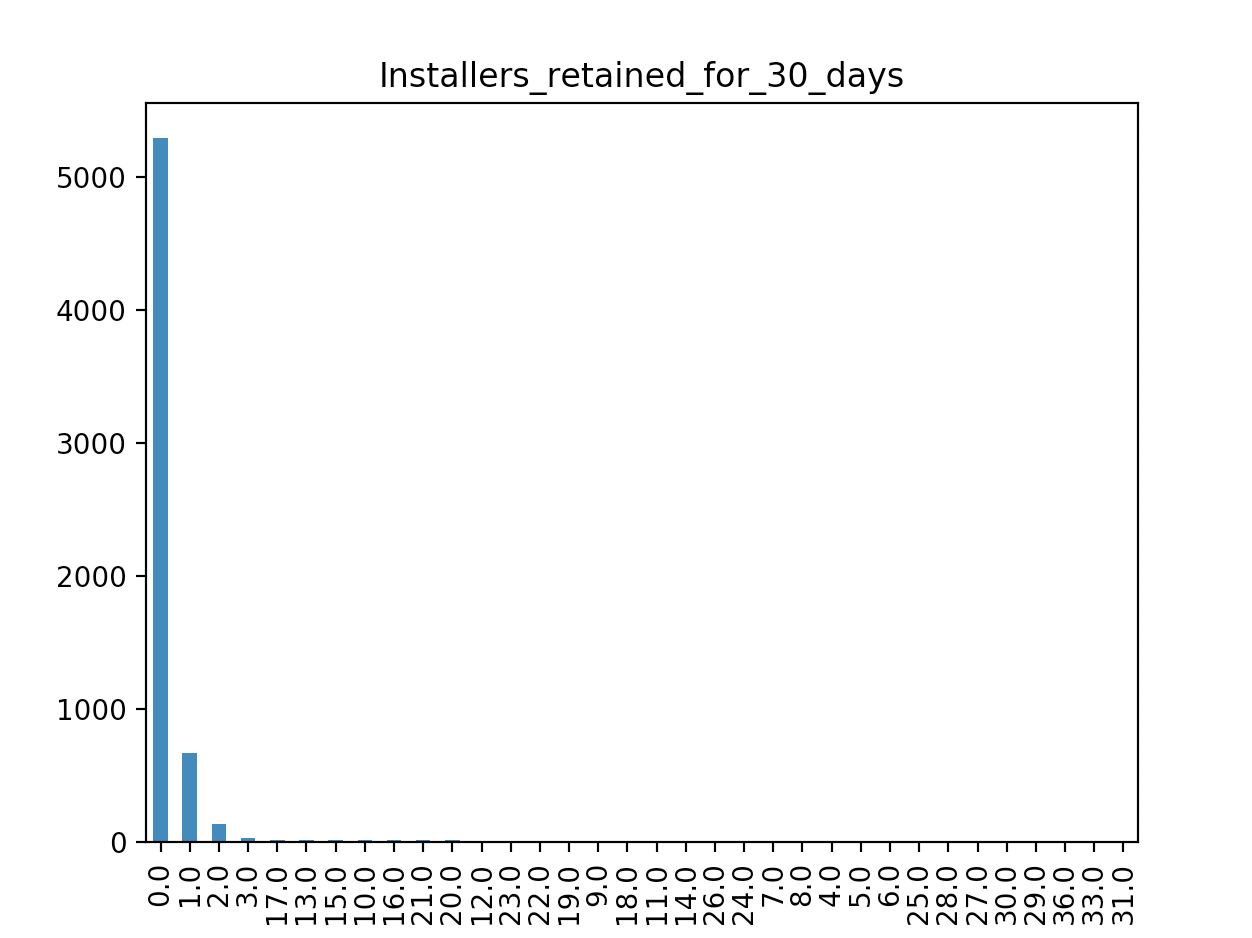
**Original data set visualizations using Python 3.7.2 and PyCharm EDU version 2019.12:**

*Linear regression using matplotlib version 3.03:*

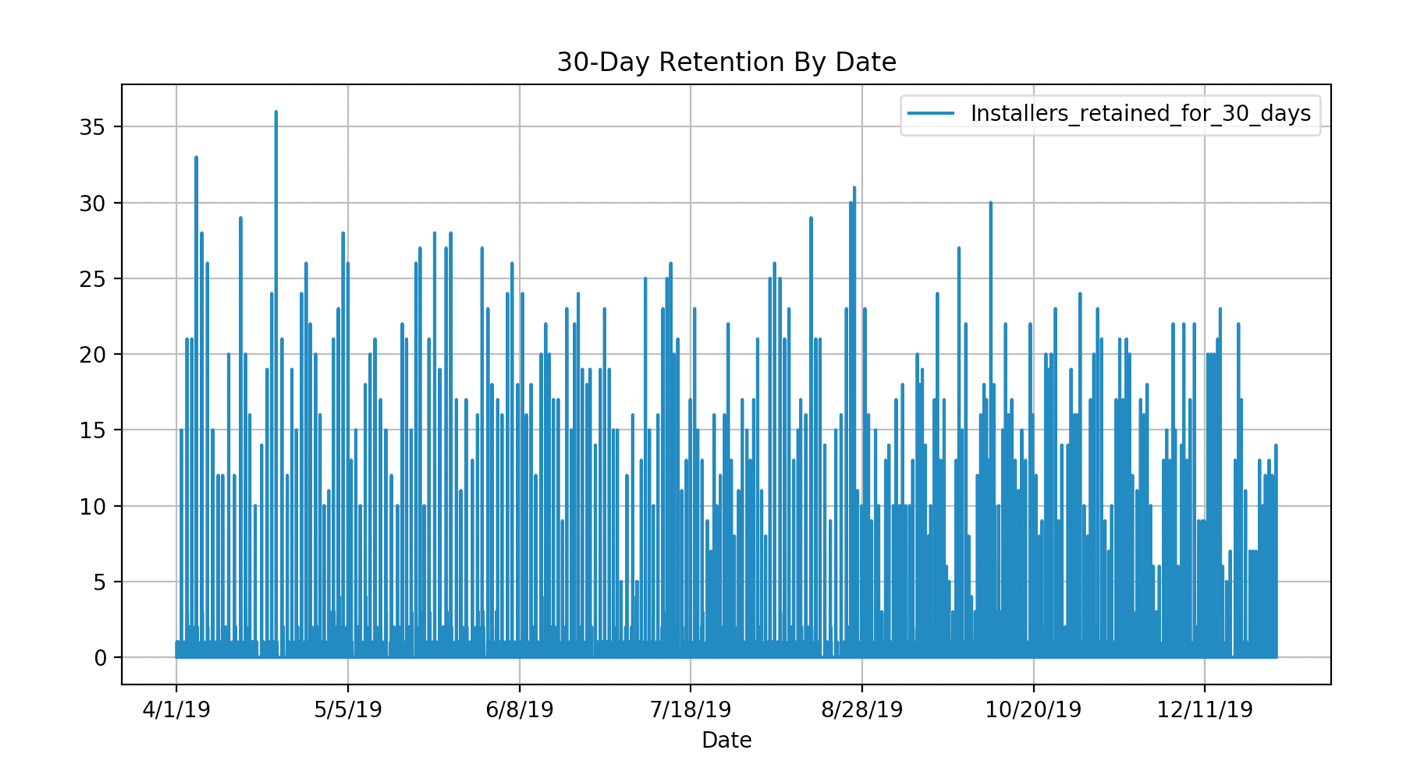
****

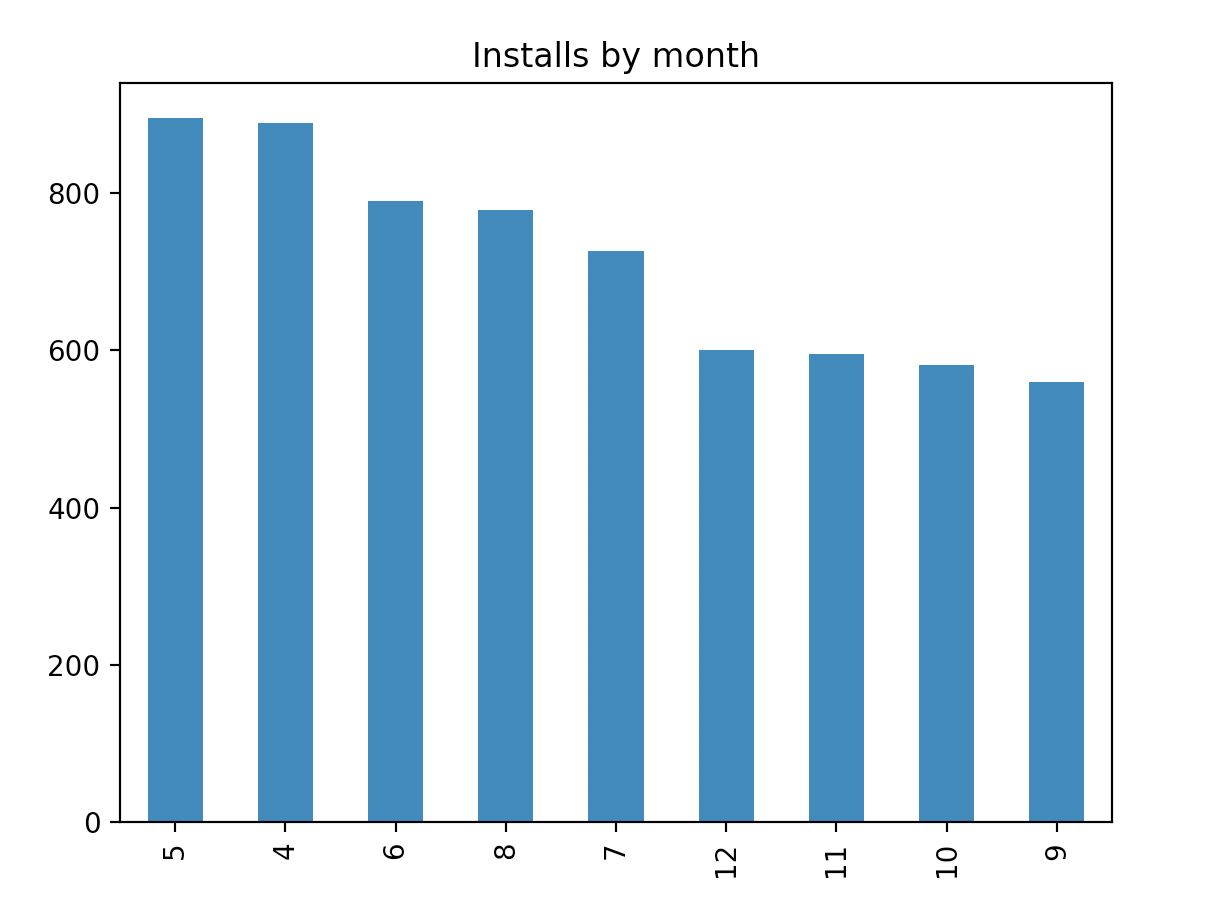
****

****

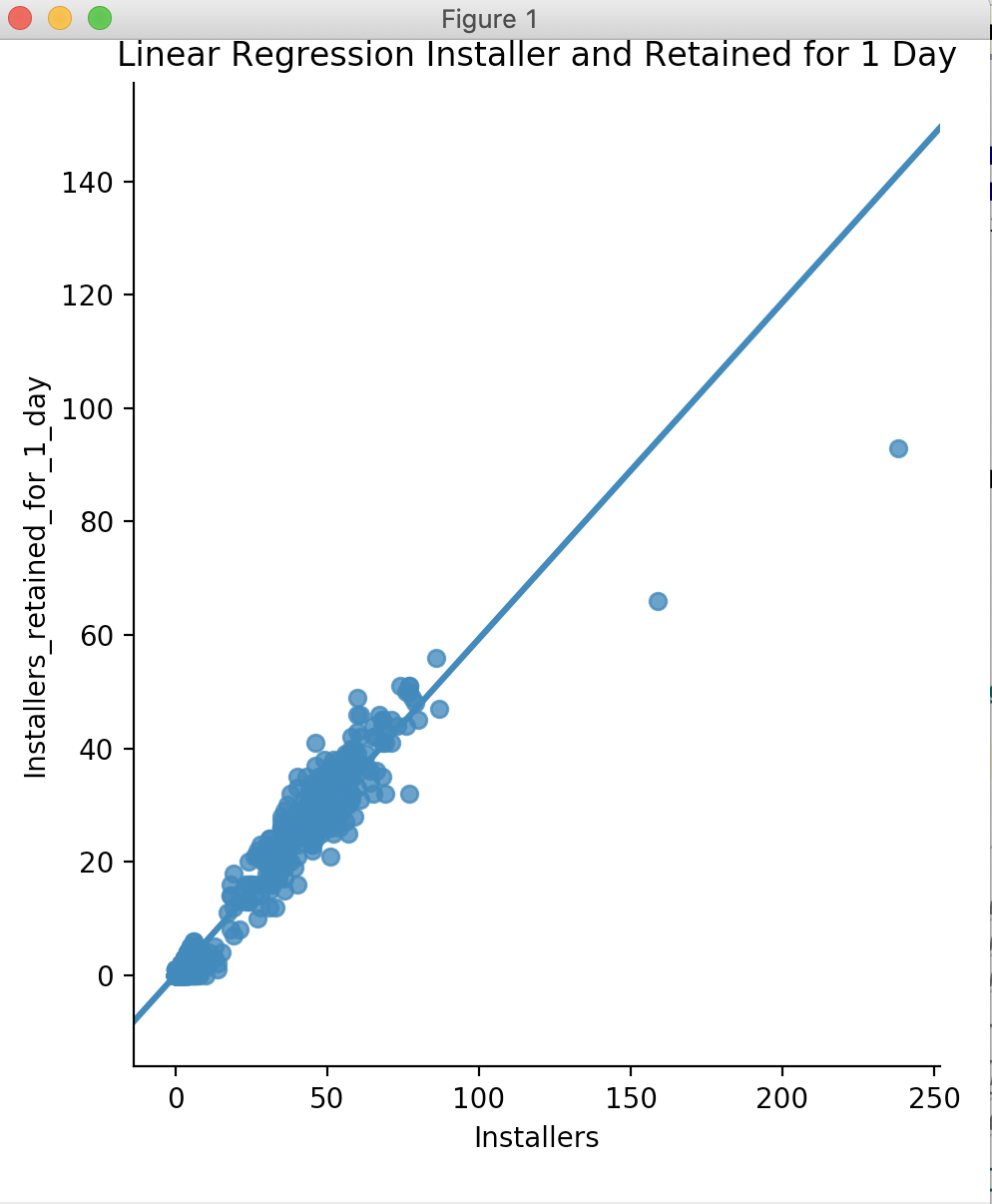
****

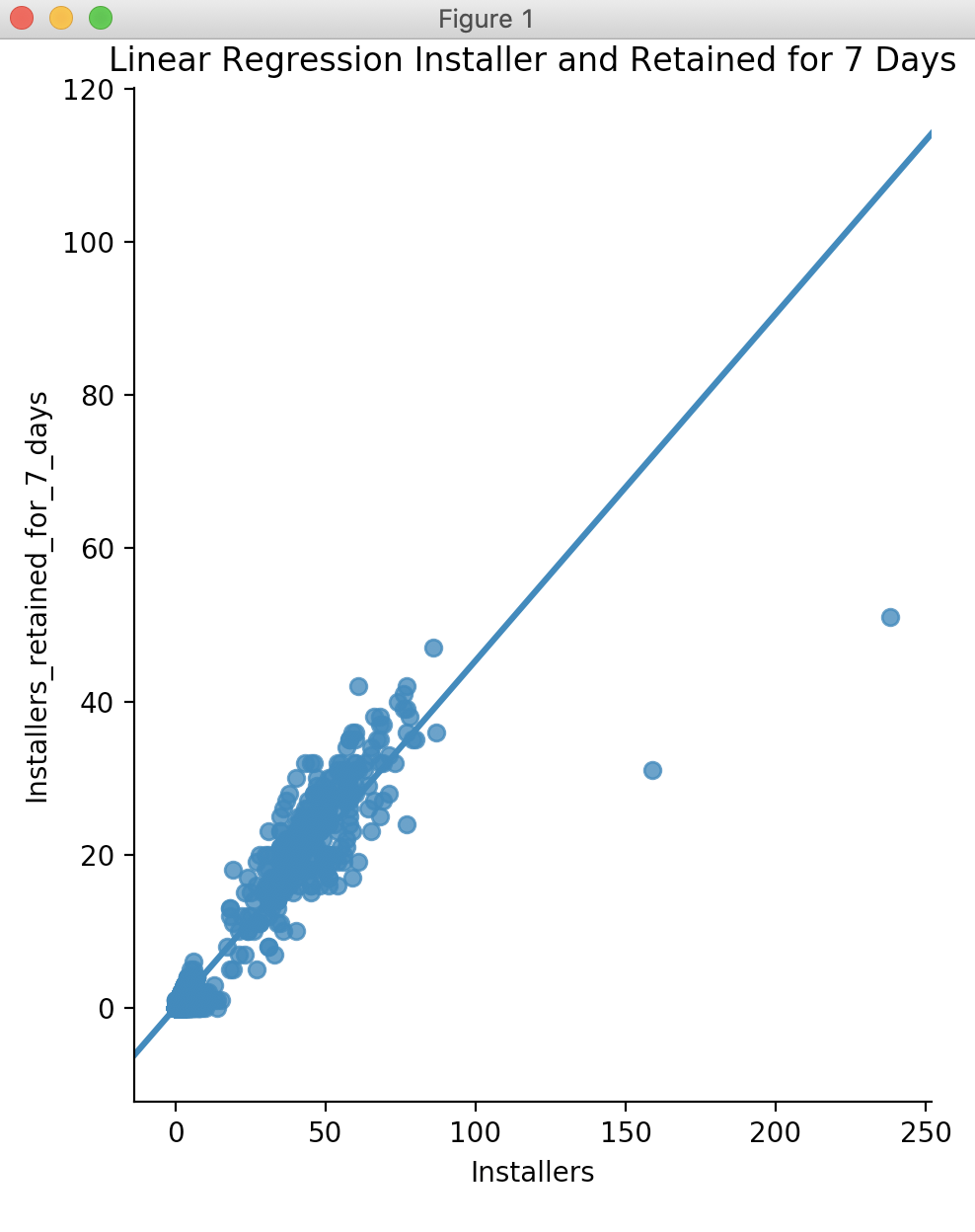
*Bar charts using matplotlib version 3.03:*

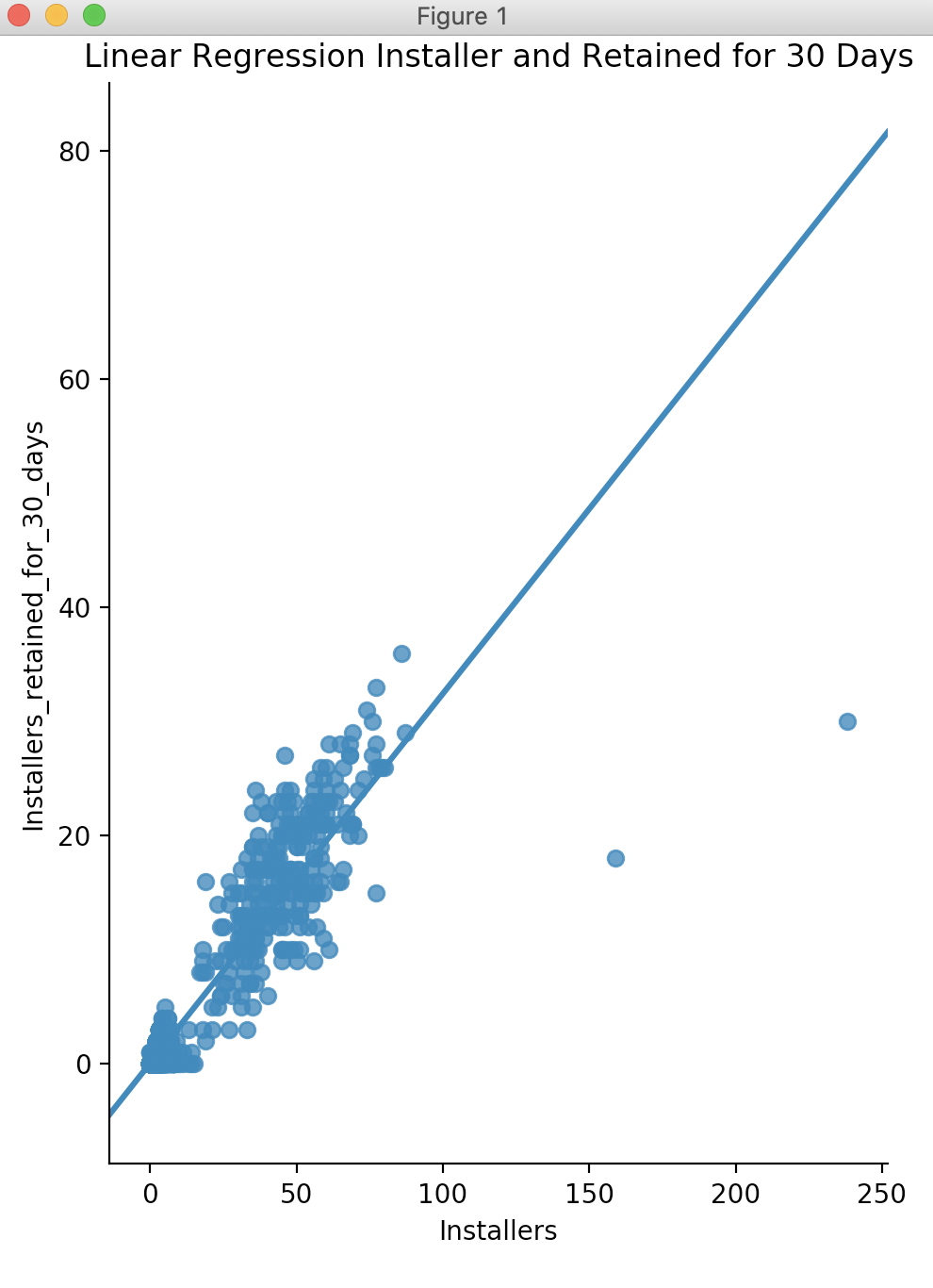
****

****

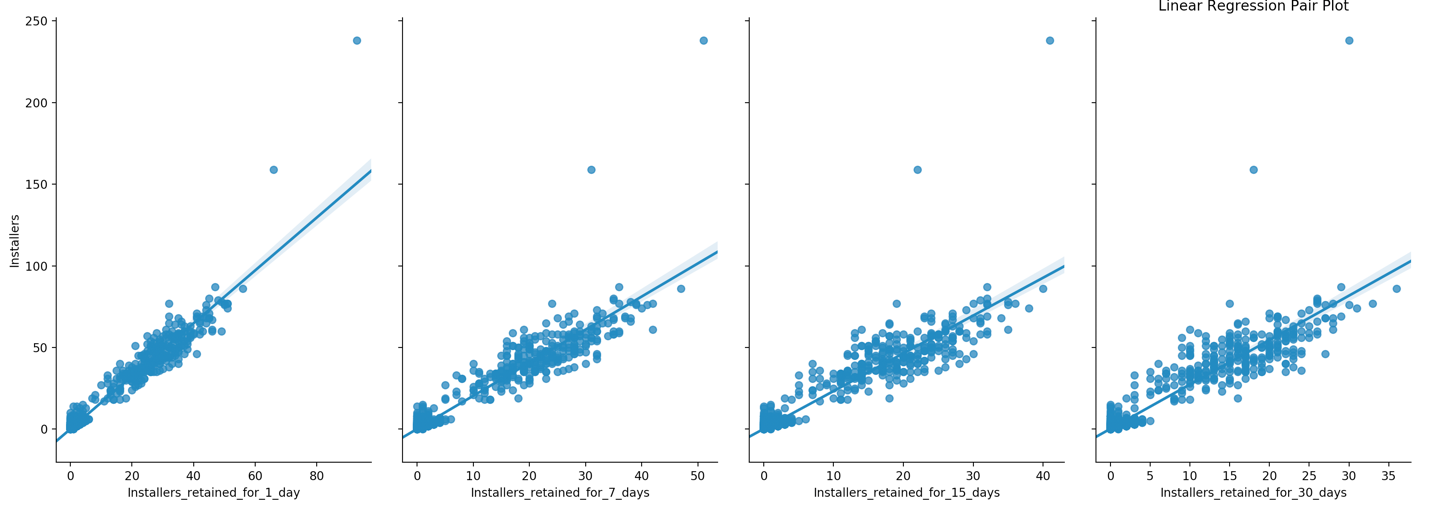
*Seaborn plots using version 0.9.0:*

****

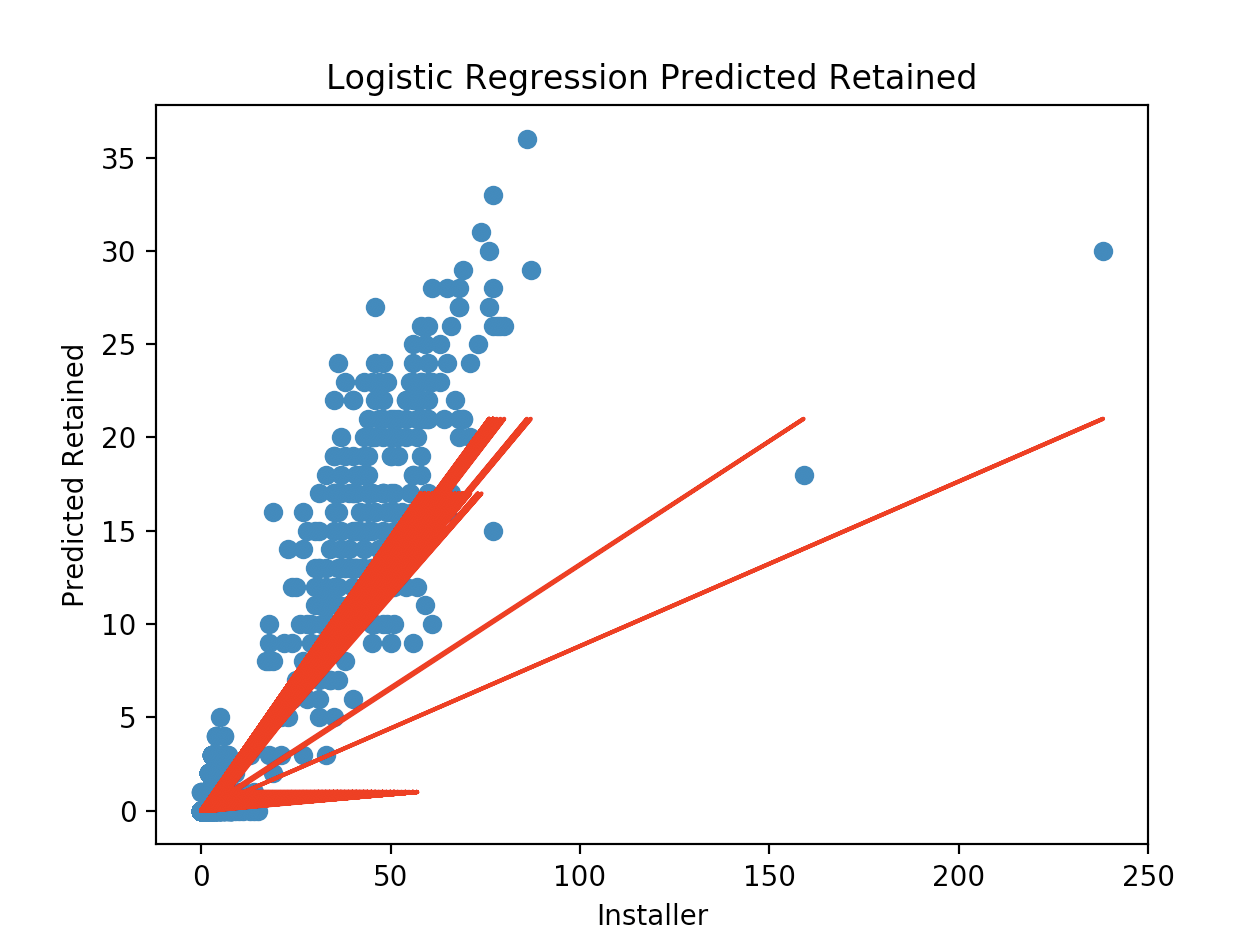
****

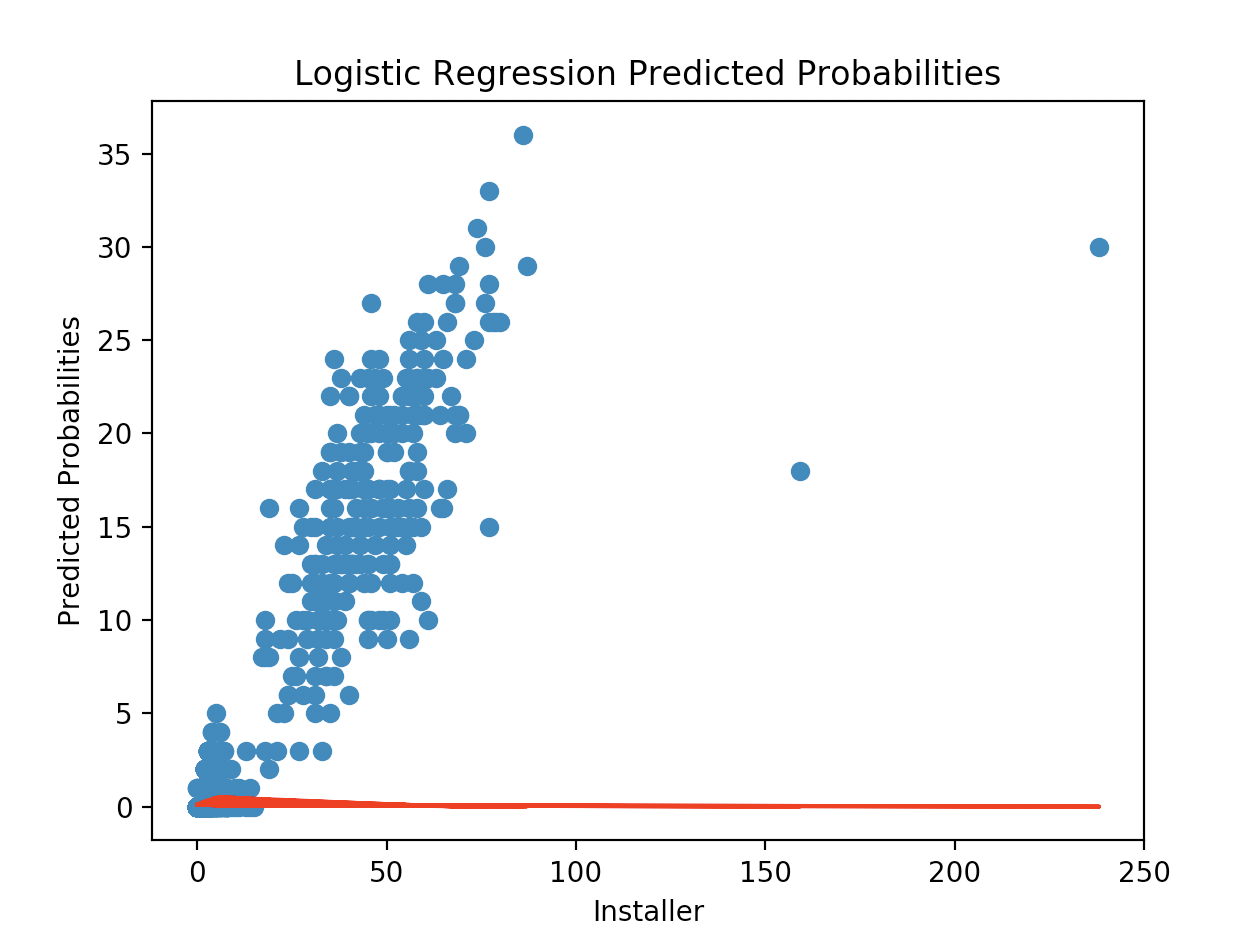
****

**After fixing missing data:**

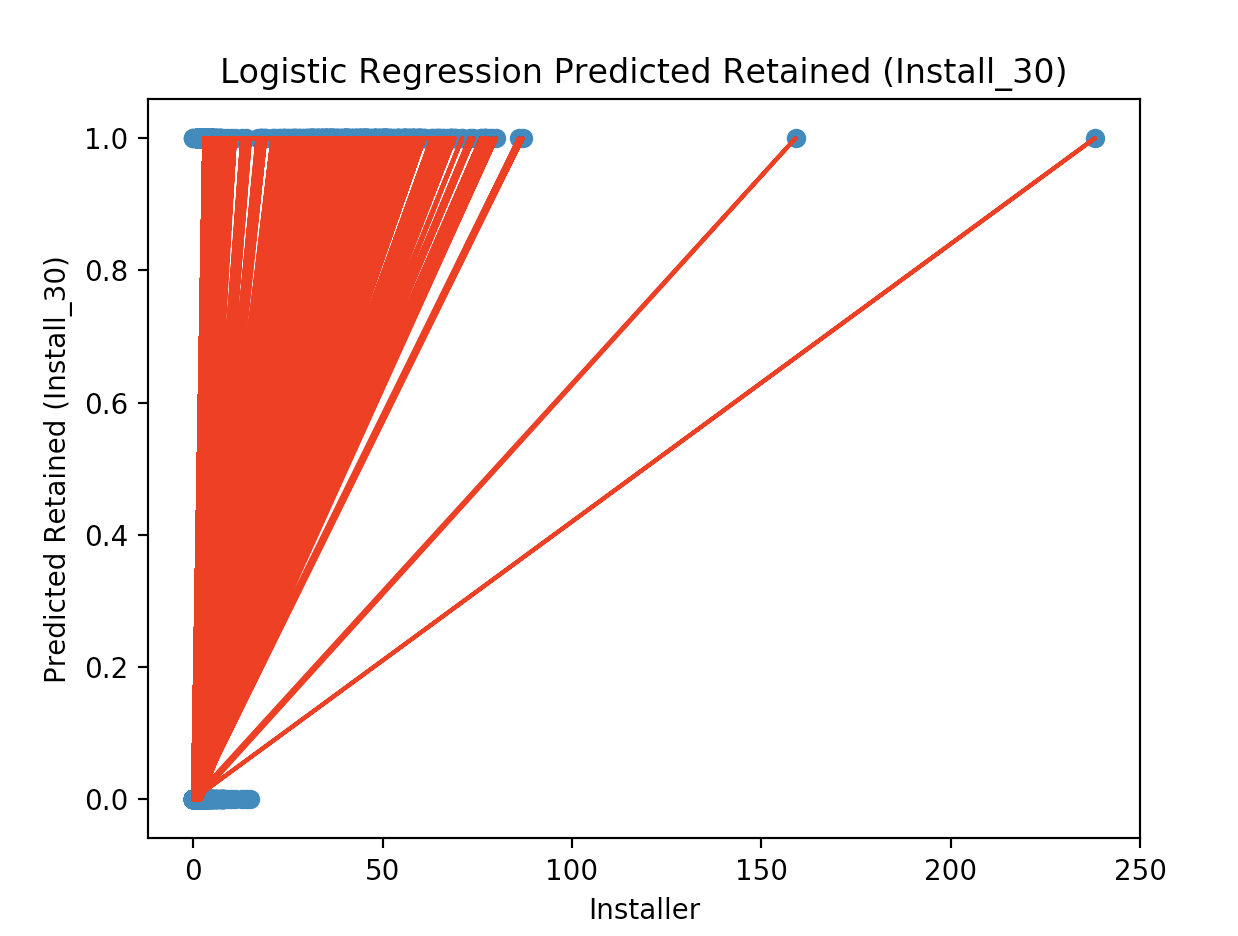
****

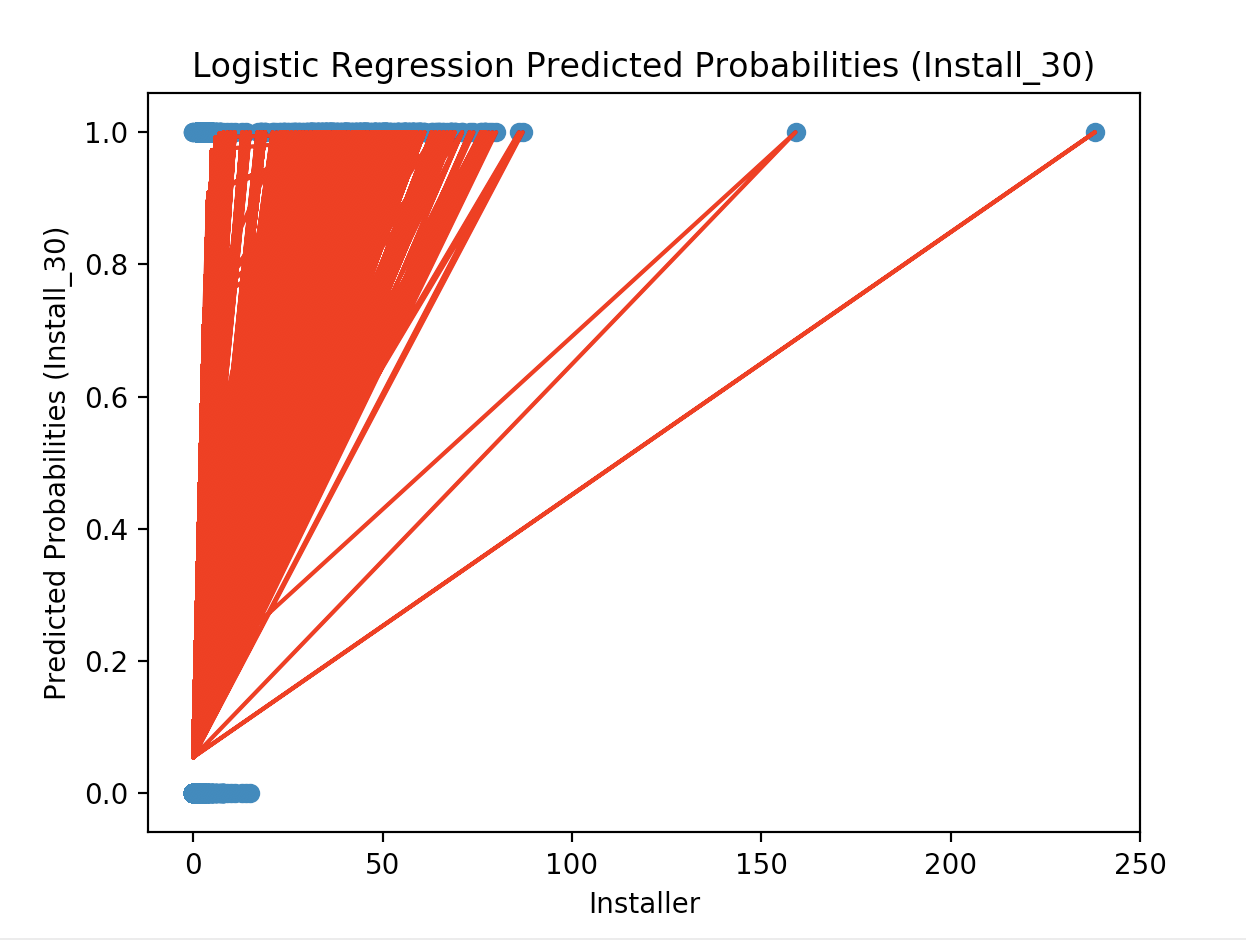
*Scatter plots using matplotlib version 3.03:*

****

****

**After adding the Install\_30 binary column for good or bad retention results:**

****

****

**Linear and Logistic Regression Results:**

Linear regression prediction for 1 install(array([0.37117343]),)

Linear regression prediction for 2 installs(array([0.69505919]),)

Linear regression prediction for 4 installs(array([1.34283072]),)

Linear regression intercept(0.04728766524424888,)

Linear regression coefficient(array([0.32388576]),)

Logistic regression probability of keeping user 30 days with 1 install (0, 1, 2 users)(array([[0.81509284, 0.10462079, 0.02104794, 0.00578331, 0.001615 ,

0.00162389, 0.00159173, 0.00169252, 0.00172722, 0.00208021,

0.00281579, 0.00195911, 0.00235012, 0.00276405, 0.00192799,

0.00235382, 0.00232595, 0.00254848, 0.00174241, 0.00194718,

0.00205494, 0.00188034, 0.00190413, 0.00190364, 0.00161867,

0.00135641, 0.00144977, 0.00134967, 0.00133698, 0.00117456,

0.00103065, 0.0011148 , 0.00111185, 0.00110325]]),)

Logistic regression probability of keeping user 30 days with 2 installs (0, 1, 2 users)(array([[0.75162555, 0.13811137, 0.0282537 , 0.00785029, 0.00213917,

0.00221759, 0.00217883, 0.00232074, 0.00236334, 0.00286865,

0.0039396 , 0.00270116, 0.00326756, 0.00388123, 0.00266191,

0.00331979, 0.00327518, 0.00358705, 0.00243522, 0.00270008,

0.00289122, 0.00266341, 0.0026564 , 0.00266567, 0.00224164,

0.00186896, 0.00202336, 0.00186038, 0.00184421, 0.00161229,

0.00142795, 0.00152121, 0.00151775, 0.00150753]]),)

Logistic regression probability of keeping user 30 days with 4 installs (0, 1, 2 users)(array([[0.32633166, 0.36121627, 0.0764096 , 0.02170917, 0.00563294,

0.00620685, 0.00612733, 0.00654867, 0.00664083, 0.00818758,

0.01157416, 0.0077068 , 0.00948044, 0.01148546, 0.00761565,

0.00991097, 0.00974633, 0.01066551, 0.00713928, 0.00779216,

0.00858971, 0.00802003, 0.00775929, 0.00784492, 0.00645235,

0.00532551, 0.00591506, 0.00530506, 0.00526642, 0.0045595 ,

0.00411388, 0.00425127, 0.00424473, 0.00422461]]),)

Probability of keeping 1 user 30 days with 1 install (Install\_30)(array([[0.82779678, 0.17220322]]),)

Probability of keeping 1 user 30 days with 2 install (Install\_30)(array([[0.57069118, 0.42930882]]),)

Probability of keeping 1 user 30 days with 4 install (Install\_30)(array([[0.09227487, 0.90772513]]),)

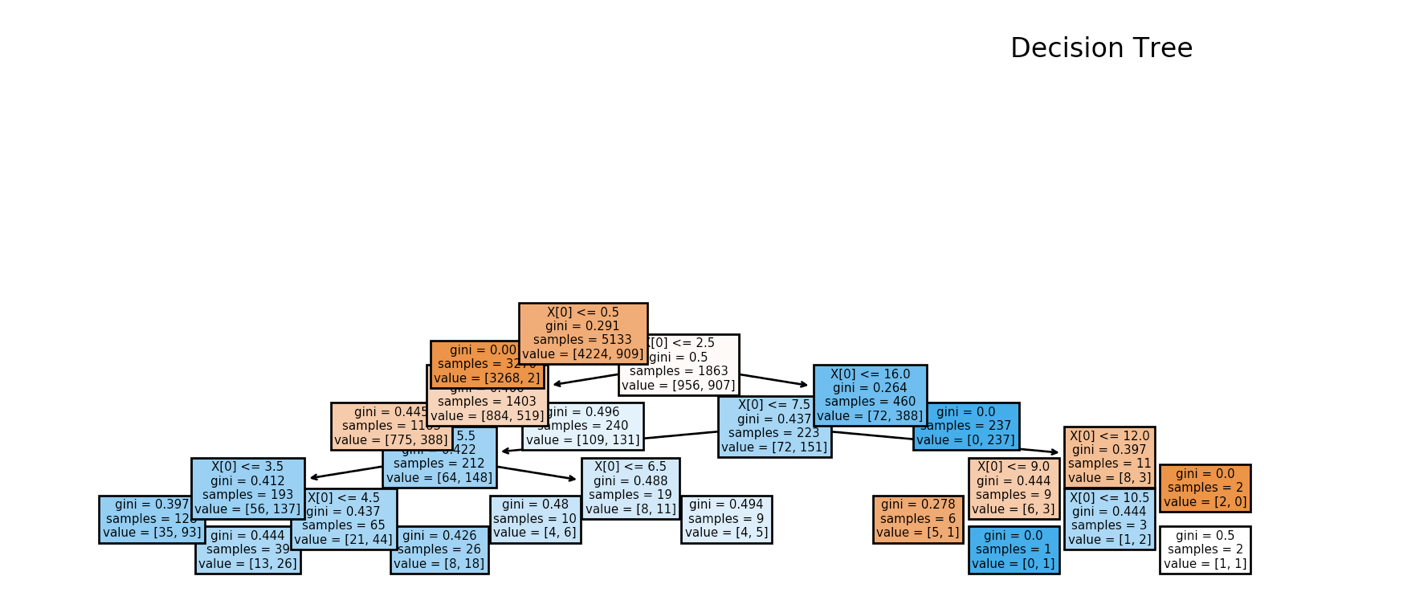
Logistic regression intercept (Install\_30)(array([-2.85551311]),)

Logistic regression coefficient (Install\_30)(array([[1.28542071]]),)

**Decision Tree and Random Forest Results:**

Decision tree model Accuracy is (0.893895065340355,)

*Decision tree plot using sklearn version 0.0:*



Accuracy for random forest (Install\_30)(0.8705148205928237,)

Random forest prediction of keeping 1 user 30 days with 1 installs (Install\_30) [0,1] is(array([0]),)

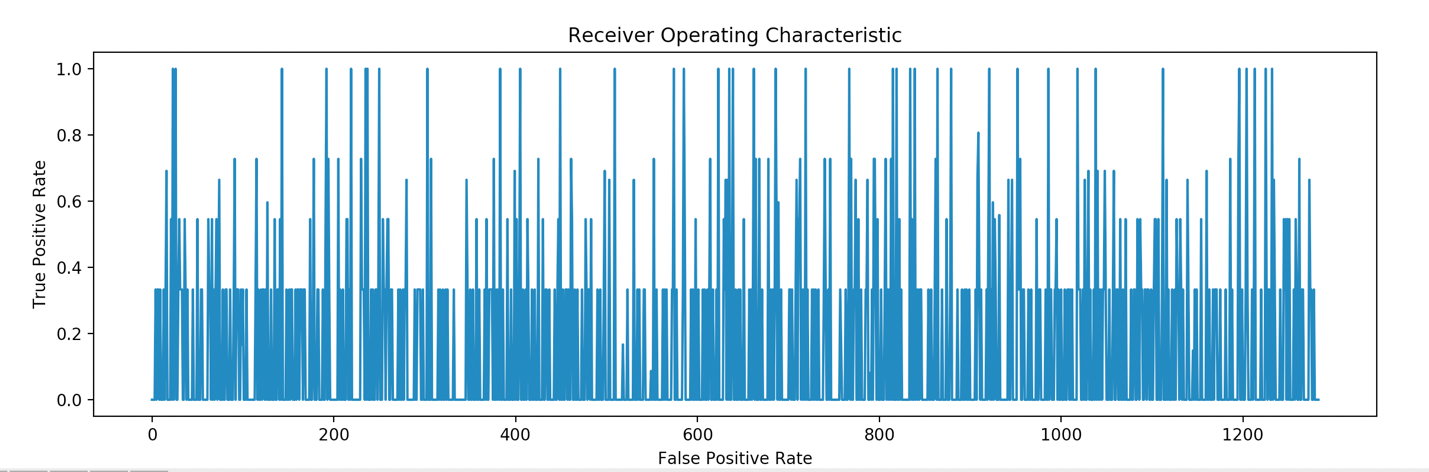
Random forest prediction of keeping 1 user 30 days with 2 installs (Install\_30) [0,1] is(array([1]),)

Random forest prediction of keeping 1 user 30 days with 4 installs (Install\_30) [0,1] is(array([1]),)

Mean Absolute Error:(0.13,)

Random Forest ROC AUC(0.9231644536522586,)

*Plot of random forest predictions using matplotlib versus 0.3.0:*



The mean absolute percentage error (MAPE) using y\_test(4588 NaN

2137 NaN

2274 0.0

1303 NaN

4576 NaN

...

3210 NaN

2483 NaN

5386 NaN

1560 NaN

3232 NaN

Name: Install\_30, Length: 1282, dtype: float64,)

Random forest accuracy percentage using error percentage(-inf,)