

# nbf\_organic

July 19, 2020

## 1 Predicting Customer Website Clicks

```
[1]: # importing libraries
from utilities import *

%matplotlib inline
%load_ext autoreload
%autoreload 2
```

Importing plotly failed. Interactive plots will not work.

```
[2]: # file paths
catalogs_file_path = r'data\NBF_Catalogs.xlsx'
organic_file_path = r"data\NBF_Web-Web_Channel_Organic_Search_20180101-20200531.
↳xlsx"
```

```
[3]: # parameters
n_forecast_weeks = 13
freq = "W"
```

```
[4]: # import catalogs data set
nbf_catalogs_ts = read_catalogs_data(catalogs_file_path, freq)
```

```
[5]: # import direct data set
nbf_organic_ts = read_clicks_data(organic_file_path, freq)
```

```
-----
['google' 'bing' 'yahoo' 'duckduckgo' 'images.google' 'ask'
 'us.search.yahoo.com' 'aol' 'ecosia.org' 'baidu' 'avg' 'google images'
 'sogou' 'yandex' 'incredimail' 'kvasir' 'msn' 'naver' 'onet' 'rakuten'
 'search-results' 'search.smt.docomo' 'seznam' 'startsiden']
```

```
[6]: # merge catalogs and direct data sets
nbf_catalogs_organic_ts = merge_catalogs_clicks(nbf_catalogs_ts, nbf_organic_ts)
```

```
-----
corr 0.055423872025481304
```

p-value 0.5376271241922508

```
-----
                        OLS Regression Results
=====
Dep. Variable:          no_clicks      R-squared:                0.003
Model:                  OLS           Adj. R-squared:          -0.005
Method:                 Least Squares  F-statistic:             0.3821
Date:                  Sun, 19 Jul 2020 Prob (F-statistic):       0.538
Time:                  19:10:33        Log-Likelihood:          -1207.3
No. Observations:      126            AIC:                    2419.
Df Residuals:          124            BIC:                    2424.
Df Model:               1
Covariance Type:       nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept      1.383e+04    538.257     25.695     0.000     1.28e+04     1.49e+04
no_catalogs      0.0015      0.002      0.618     0.538     -0.003      0.006
=====
Omnibus:                 62.222    Durbin-Watson:           0.167
Prob(Omnibus):           0.000    Jarque-Bera (JB):         8.278
Skew:                    0.039    Prob(JB):                 0.0159
Kurtosis:                1.747    Cond. No.                  3.69e+05
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.69e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
[7]: # make predictions
clicks_ts(nbf_catalogs_organic_ts, n_forecast_weeks, freq)
```

```
-----
      ds      yhat    yhat_lower    yhat_upper
126 2020-06-07  8794.782327  6600.213583  11048.609551
127 2020-06-14  8537.313807  6208.905698  10784.531141
128 2020-06-21  8264.485085  5939.522801  10332.761757
129 2020-06-28  7940.198491  5791.391453  10165.163723
130 2020-07-05  7813.725944  5401.885827  10083.601958
131 2020-07-12  8131.191970  5898.161990  10466.343331
132 2020-07-19  8770.866722  6461.462365  11070.140559
133 2020-07-26  9269.561653  6886.924085  11569.028145
134 2020-08-02  9237.254498  7033.380044  11565.225362
135 2020-08-09  8698.758920  6470.892210  11018.156728
136 2020-08-16  8010.834669  5865.213028  10332.289544
137 2020-08-23  7504.942302  5068.750215  9738.330880
```

138 2020-08-30 7266.217737 4969.853241 9325.224528

