ADS505_Final

October 7, 2022

1 Predicting Bike Rental Counts in Seoul Based on the Weather and Holiday Information for a Stable Supply

2 Team 5

2.1 Kyle Dalope

3 ADS505-01-FA22

[2]:

```
! pip install dmba
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting dmba
  Downloading dmba-0.1.0-py3-none-any.whl (11.8 MB)
     |                  | 11.8 MB 11.6 MB/s
Installing collected packages: dmba
Successfully installed dmba-0.1.0
```

[3]:

```
#Imports Required

import pandas as pd
import matplotlib.pylab as plt
import numpy as np

from sklearn import preprocessing
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from dmba import gainsChart, liftChart
from dmba import classificationSummary
from sklearn.metrics import f1_score, precision_score, recall_score,␣
 ↪accuracy_score

%matplotlib inline
```

```
no display found. Using non-interactive Agg backend
```

4 EDA

[4]:

```
#Data import

SeoulBike_df = pd.read_csv("SeoulBikeData.csv", encoding = 'unicode_escape')
```

[5]:

```
SeoulBike_df.head()
```

[5]:

```
         Date  Rented Bike Count  Hour  Temperature(°C)  Humidity(%)  \
0  01/12/2017                254     0             -5.2           37
1  01/12/2017                204     1             -5.5           38
2  01/12/2017                173     2             -6.0           39
3  01/12/2017                107     3             -6.2           40
4  01/12/2017                 78     4             -6.0           36

   Wind speed (m/s)  Visibility (10m)  Dew point temperature(°C)  \
0               2.2              2000                      -17.6
1               0.8              2000                      -17.6
2               1.0              2000                      -17.7
3               0.9              2000                      -17.6
4               2.3              2000                      -18.6

   Solar Radiation (MJ/m2)  Rainfall(mm)  Snowfall (cm) Seasons     Holiday  \
0                      0.0           0.0            0.0  Winter  No Holiday
1                      0.0           0.0            0.0  Winter  No Holiday
2                      0.0           0.0            0.0  Winter  No Holiday
3                      0.0           0.0            0.0  Winter  No Holiday
4                      0.0           0.0            0.0  Winter  No Holiday

  Functioning Day
0             Yes
1             Yes
2             Yes
3             Yes
4             Yes
```

[6]:

```
SeoulBike_df.describe() #Statistical summary
```

[6]:

```
       Rented Bike Count         Hour  Temperature(°C)  Humidity(%)  \
count        8760.000000  8760.000000      8760.000000  8760.000000
mean          704.602055    11.500000        12.882922    58.226256
std           644.997468     6.922582        11.944825    20.362413
min             0.000000     0.000000       -17.800000     0.000000
25%           191.000000     5.750000         3.500000    42.000000
50%           504.500000    11.500000        13.700000    57.000000
75%          1065.250000    17.250000        22.500000    74.000000
max          3556.000000    23.000000        39.400000    98.000000

       Wind speed (m/s)  Visibility (10m)  Dew point temperature(°C)  \
count       8760.000000       8760.000000                8760.000000
mean           1.724909       1436.825799                   4.073813
std            1.036300        608.298712                  13.060369
min            0.000000         27.000000                 -30.600000
25%            0.900000        940.000000                  -4.700000
50%            1.500000       1698.000000                   5.100000
75%            2.300000       2000.000000                  14.800000
max            7.400000       2000.000000                  27.200000

       Solar Radiation (MJ/m2)  Rainfall(mm)  Snowfall (cm)
count              8760.000000   8760.000000    8760.000000
mean                  0.569111      0.148687       0.075068
std                   0.868746      1.128193       0.436746
min                   0.000000      0.000000       0.000000
25%                   0.000000      0.000000       0.000000
50%                   0.010000      0.000000       0.000000
75%                   0.930000      0.000000       0.000000
max                   3.520000     35.000000       8.800000
```

[7]:

```
SeoulBike_df.info() #Observe data types
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8760 entries, 0 to 8759
Data columns (total 14 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Date                       8760 non-null   object
 1   Rented Bike Count          8760 non-null   int64
 2   Hour                       8760 non-null   int64
 3   Temperature(°C)            8760 non-null   float64
 4   Humidity(%)                8760 non-null   int64
 5   Wind speed (m/s)           8760 non-null   float64
 6   Visibility (10m)           8760 non-null   int64
 7   Dew point temperature(°C)  8760 non-null   float64
 8   Solar Radiation (MJ/m2)    8760 non-null   float64
 9   Rainfall(mm)               8760 non-null   float64
 10   Snowfall (cm)              8760 non-null   float64
 11   Seasons                    8760 non-null   object
 12   Holiday                    8760 non-null   object
 13   Functioning Day            8760 non-null   object
dtypes: float64(6), int64(4), object(4)
memory usage: 958.2+ KB
```

[8]:

```
SeoulBike_df.isnull().sum() #Observe if any missing data exists
```

[8]:

```
Date                         0
Rented Bike Count            0
Hour                         0
Temperature(°C)              0
Humidity(%)                  0
Wind speed (m/s)             0
Visibility (10m)             0
Dew point temperature(°C)    0
Solar Radiation (MJ/m2)      0
Rainfall(mm)                 0
Snowfall (cm)                0
Seasons                      0
Holiday                      0
Functioning Day              0
dtype: int64
```

5 Data Pre-processing

[9]:

```
#Reformat Column Names

SeoulBike_df = SeoulBike_df.copy()

SeoulBike_df.columns = [d.replace(' ', '_').replace('.', '') for d in␣
 ↪SeoulBike_df.columns]
SeoulBike_df.head()
```

[9]:

```
         Date  Rented_Bike_Count  Hour  Temperature(°C)  Humidity(%)  \
0  01/12/2017                254     0             -5.2           37
1  01/12/2017                204     1             -5.5           38
2  01/12/2017                173     2             -6.0           39
3  01/12/2017                107     3             -6.2           40
4  01/12/2017                 78     4             -6.0           36

   Wind_speed_(m/s)  Visibility_(10m)  Dew_point_temperature(°C)  \
0               2.2              2000                      -17.6
1               0.8              2000                      -17.6
```

```
2                1.0              2000                         -17.7
3                0.9              2000                         -17.6
4                2.3              2000                         -18.6

   Solar_Radiation_(MJ/m2)  Rainfall(mm)  Snowfall_(cm) Seasons     Holiday  \
0                      0.0           0.0            0.0  Winter  No Holiday
1                      0.0           0.0            0.0  Winter  No Holiday
2                      0.0           0.0            0.0  Winter  No Holiday
3                      0.0           0.0            0.0  Winter  No Holiday
4                      0.0           0.0            0.0  Winter  No Holiday

  Functioning_Day
0             Yes
1             Yes
2             Yes
3             Yes
4             Yes
```

[12]:

```
SeoulBike_df1 = pd.get_dummies(SeoulBike_df[['Seasons', 'Holiday',␣
 ↪'Fuctioning_Day']], prefix_sep='_')
SeoulBike_df1
```

```
      ␣
↪---------------------------------------------------------------------------

       KeyError                                  Traceback (most recent call␣
↪last)

       <ipython-input-12-093c90fb9513> in <module>
   ----> 1 SeoulBike_df1 = pd.get_dummies(SeoulBike_df[['Seasons', 'Holiday',␣
↪'Fuctioning_Day']], prefix_sep='_')
         2 SeoulBike_df1


       /usr/local/lib/python3.7/dist-packages/pandas/core/frame.py in␣
↪__getitem__(self, key)
      3462             if is_iterator(key):
      3463                 key = list(key)
   -> 3464             indexer = self.loc._get_listlike_indexer(key, axis=1)[1]
      3465
      3466         # take() does not accept boolean indexers


       /usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py in␣
↪_get_listlike_indexer(self, key, axis)
```

```
    1312             keyarr, indexer, new_indexer = ax.
↪_reindex_non_unique(keyarr)
    1313 
 -> 1314         self._validate_read_indexer(keyarr, indexer, axis)
    1315 
    1316         if needs_i8_conversion(ax.dtype) or isinstance(


     /usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py in␣
↪_validate_read_indexer(self, key, indexer, axis)
    1375 
    1376             not_found = list(ensure_index(key)[missing_mask.
↪nonzero()[0]].unique())
 -> 1377             raise KeyError(f"{not_found} not in index")
    1378 
    1379 


     KeyError: "['Fuctioning_Day'] not in index"
```

[ ]:

6 Model Selections

[ ]:

[ ]:

[ ]:

7 Model Evaluation

[ ]:

[ ]:

[ ]:

8 Final Model Selection and Conclusion

[ ]:

```

```

[ ]:

```
!sudo apt-get install texlive-xetex texlive-fonts-recommended texlive-plain-generic
!jupyter nbconvert --to pdf /content/ADS505_Final.ipynb
```