Cycle Data clean part 01-DBConnect

April 27, 2018

1 6.4 Cycle Data Munging

Declaration : The coding is used was abstract from Kevin mark ham youtube video seriese, Introduction to machine learning with scikit-learn video series. You can find link under resources section.

2 6.4.1 Data Munging steps

1. Handling missing Data
2. Encoding, Decoding, and recoding of data
3. Handling Anomalous values
4. Transforming Data
5. Merging Dataset

3 Data Cleaning

```
In [2]: # load libraries and set styles, options
        import os,csv
        import numpy as np
        import pandas as pd
        import seaborn as sns
        from IPython.display import HTML
        import warnings; warnings.simplefilter('ignore')
```

```
In [3]: %matplotlib inline
```

2. Read and verify data

```
In [4]: # read in a CSV
        df = pd.read_csv('C:/Users/mrferozi/Desktop/GitHub/Bike/cycle-share-dataset-orignal/tr
```

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 286857 entries, 0 to 286856
Data columns (total 12 columns):
trip_id              286857 non-null int64
```

```
starttime            286857 non-null object
stoptime             286857 non-null object
bikeid               286857 non-null object
tripduration         286857 non-null float64
from_station_name    286857 non-null object
to_station_name      286857 non-null object
from_station_id      286857 non-null object
to_station_id        286857 non-null object
usertype             286857 non-null object
gender               181557 non-null object
birthyear            181553 non-null float64
dtypes: float64(2), int64(1), object(9)
memory usage: 26.3+ MB
```

The data set contains trip duration for less than three minutes, which could be possible because it may happen that the user took the bike from cycle dock and decide not to proceed with the journey and the user return bike to the dock. Another, thing which we are ignoring for the moment is the journey recorded for the short pass holder which does not kick in any meaning in our current analysis, Note, a separate analysis has performed for short pass holders but at this point this study decide to get rid of undesirable data and more focus on the data which contributing significance in our analysis.

```
In [5]: # Filter data and only consider trip duration more than 3 minute and exclude Short-Ter
        bikes = df[(df.tripduration >=300) & (df.usertype != 'Short-Term Pass Holder')]
```

```
In [6]: bikes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 143543 entries, 0 to 286847
Data columns (total 12 columns):
trip_id              143543 non-null int64
starttime            143543 non-null object
stoptime             143543 non-null object
bikeid               143543 non-null object
tripduration         143543 non-null float64
from_station_name    143543 non-null object
to_station_name      143543 non-null object
from_station_id      143543 non-null object
to_station_id        143543 non-null object
usertype             143543 non-null object
gender               143543 non-null object
birthyear            143541 non-null float64
dtypes: float64(2), int64(1), object(9)
memory usage: 14.2+ MB
```

- After applying fillter our data set shrink from 286857 to 143541 tuples

4 Imputation (Handle missing values)

What does "NaN" mean?

- "NaN" is not a string, rather it's a special value: `numpy.nan`.
- It stands for "Not a Number" and indicates a **missing value**.
- `read_csv` detects missing values (by default) when reading the file, and replaces them with this special value.

```
In [7]: # 'isnull' returns a DataFrame of booleans (True if missing, False if not missing)
        bikes.isnull().tail()

Out[7]:        trip_id starttime stoptime bikeid tripduration from_station_name  \
        286842   False     False    False  False        False             False
        286843   False     False    False  False        False             False
        286844   False     False    False  False        False             False
        286845   False     False    False  False        False             False
        286847   False     False    False  False        False             False

               to_station_name from_station_id to_station_id usertype gender birthyear
        286842           False           False         False    False  False     False
        286843           False           False         False    False  False     False
        286844           False           False         False    False  False     False
        286845           False           False         False    False  False     False
        286847           False           False         False    False  False     False

In [8]: # 'nonnull' returns the opposite of 'isnull' (True if not missing, False if missing)
        bikes.notnull().tail()

Out[8]:        trip_id starttime stoptime bikeid tripduration from_station_name  \
        286842    True      True     True   True         True              True
        286843    True      True     True   True         True              True
        286844    True      True     True   True         True              True
        286845    True      True     True   True         True              True
        286847    True      True     True   True         True              True

               to_station_name from_station_id to_station_id usertype gender birthyear
        286842            True            True          True     True   True      True
        286843            True            True          True     True   True      True
        286844            True            True          True     True   True      True
        286845            True            True          True     True   True      True
        286847            True            True          True     True   True      True

In [9]: # count the number of missing values in each Series
        bikes.isnull().sum()

Out[9]: trip_id                0
        starttime              0
        stoptime               0
```

```
bikeid               0
tripduration         0
from_station_name    0
to_station_name      0
from_station_id      0
to_station_id        0
usertype             0
gender               0
birthyear            2
dtype: int64
```

handle missing values depends on the dataset as well as the nature of analysis

```
In [10]: bikes.shape

Out[10]: (143543, 12)

In [11]: # if 'any' values are missing in a row, then drop that row
         bikes.dropna(how='any').shape

Out[11]: (143541, 12)

In [12]: bikes.birthyear.mean()

Out[12]: 1979.3765474672741

In [13]: # fill in missing values with a specified value
         bikes['birthyear'].fillna(value='1979', inplace=True)

In [14]: # count the number of missing values in each Series
         bikes.isnull().sum()

Out[14]: trip_id              0
         starttime            0
         stoptime             0
         bikeid               0
         tripduration         0
         from_station_name    0
         to_station_name      0
         from_station_id      0
         to_station_id        0
         usertype             0
         gender               0
         birthyear            0
         dtype: int64
```

5 Data Encoding

converting data Categorical to Numaric for further use

- Converting Gender Column

```
In [15]: # create the 'Sex_num' dummy variable using the 'map' method
         bikes['Sex_num'] = bikes.gender.map({'Female':2, 'Male':1})
```

- Converting category From Station ID and To Station ID to nominal

```
In [16]: bikes["from_station_id_cat"] = bikes["from_station_id"].astype('category')
         bikes.dtypes
```

```
Out[16]: trip_id                   int64
         starttime                object
         stoptime                 object
         bikeid                   object
         tripduration            float64
         from_station_name        object
         to_station_name          object
         from_station_id          object
         to_station_id            object
         usertype                 object
         gender                   object
         birthyear                object
         Sex_num                 float64
         from_station_id_cat    category
         dtype: object
```

```
In [17]: bikes["from_station_id_num"] = bikes["from_station_id_cat"].cat.codes
```

```
In [18]: bikes["to_station_id_cat"] = bikes["to_station_id"].astype('category')
         bikes.dtypes
```

```
Out[18]: trip_id                   int64
         starttime                object
         stoptime                 object
         bikeid                   object
         tripduration            float64
         from_station_name        object
         to_station_name          object
         from_station_id          object
         to_station_id            object
         usertype                 object
         gender                   object
         birthyear                object
         Sex_num                 float64
         from_station_id_cat    category
         from_station_id_num        int8
         to_station_id_cat      category
         dtype: object
```

```
In [19]: bikes["to_station_id_num"] = bikes["to_station_id_cat"].cat.codes
```

6 Data Encoding

For further, for our analysis we are dividing date columns in small portion. Which, could help us to understand dataset more clearly.

```
In [20]: # convert 'Time' to datetime format
         bikes['starttime'] = pd.to_datetime(bikes.starttime)

In [21]: bikes.dtypes

Out[21]: trip_id                          int64
         starttime               datetime64[ns]
         stoptime                        object
         bikeid                          object
         tripduration                   float64
         from_station_name               object
         to_station_name                 object
         from_station_id                 object
         to_station_id                   object
         usertype                        object
         gender                          object
         birthyear                       object
         Sex_num                        float64
         from_station_id_cat           category
         from_station_id_num               int8
         to_station_id_cat             category
         to_station_id_num                 int8
         dtype: object

In [22]: bikes['Day'] = bikes.starttime.dt.weekday_name

In [23]: # convert 'Time' to datetime format
         bikes['stoptime'] = pd.to_datetime(bikes.stoptime)

In [24]: bikes["Day_cat"] = bikes["Day"].astype('category')
         bikes.dtypes

Out[24]: trip_id                          int64
         starttime               datetime64[ns]
         stoptime                datetime64[ns]
         bikeid                          object
         tripduration                   float64
         from_station_name               object
         to_station_name                 object
         from_station_id                 object
         to_station_id                   object
         usertype                        object
         gender                          object
         birthyear                       object
```

```
        Sex_num                          float64
        from_station_id_cat             category
        from_station_id_num                 int8
        to_station_id_cat               category
        to_station_id_num                   int8
        Day                               object
        Day_cat                         category
        dtype: object
```

```
In [25]: bikes["Day_num"] = bikes["Day_cat"].cat.codes
```

```
In [26]: # convenient Series attributes are now available
         bikes["sthours"] = bikes.starttime.dt.hour
```

```
In [27]: # convenient Series attributes are now available
         bikes["stphours"] = bikes.stoptime.dt.hour
```

```
In [28]: bikes['tripduration_minutes']=bikes['tripduration']/60
         desred_decimals = 2
         bikes['tripduration_minutes'] = bikes['tripduration_minutes'].apply(lambda x: round(x
```

```
In [29]:  bikes['birthyear'] = bikes['birthyear'].astype(int)
```

```
In [30]: bikes['age'] = 2018 - bikes['birthyear']
         desred_decimals = 2
         bikes['birthyear'] = bikes['birthyear'].apply(lambda x: round(x,desred_decimals))
         bikes.age.head()
```

```
Out[30]: 0    58
         1    48
         2    30
         3    41
         4    47
         Name: age, dtype: int32
```

Some statistics about age Variable

```
In [31]: bikes.age.describe()
```

```
Out[31]: count    143543.000000
         mean         38.623458
         std          10.251359
         min          19.000000
         25%          31.000000
         50%          36.000000
         75%          44.000000
         max          87.000000
         Name: age, dtype: float64
```

```
In [32]: bikes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 143543 entries, 0 to 286847
Data columns (total 24 columns):
trip_id                 143543 non-null int64
starttime               143543 non-null datetime64[ns]
stoptime                143543 non-null datetime64[ns]
bikeid                  143543 non-null object
tripduration            143543 non-null float64
from_station_name       143543 non-null object
to_station_name         143543 non-null object
from_station_id         143543 non-null object
to_station_id           143543 non-null object
usertype                143543 non-null object
gender                  143543 non-null object
birthyear               143543 non-null int64
Sex_num                 140905 non-null float64
from_station_id_cat     143543 non-null category
from_station_id_num     143543 non-null int8
to_station_id_cat       143543 non-null category
to_station_id_num       143543 non-null int8
Day                     143543 non-null object
Day_cat                 143543 non-null category
Day_num                 143543 non-null int8
sthours                 143543 non-null int64
stphours                143543 non-null int64
tripduration_minutes    143543 non-null float64
age                     143543 non-null int32
dtypes: category(3), datetime64[ns](2), float64(3), int32(1), int64(4), int8(3), object(8)
memory usage: 21.1+ MB
```

7 Removing Duplicate values

```
In [33]: # detect duplicate trip_id codes: True if an item is identical to a previous item
         bikes.trip_id.duplicated().tail()
```

```
Out[33]: 286842    False
         286843    False
         286844    False
         286845    False
         286847    False
         Name: trip_id, dtype: bool
```

The above codes shows that we do not have any duplicated rows in our dataset

```
In [34]: # Breaking date column for further Analysis
         bikes['bmonth'] = bikes['starttime'].dt.month
```

```
In [35]: # Breaking date column for further Analysis
         bikes['Date'] = bikes['starttime'].dt.date
```

```
In [36]: bikes.dtypes
```

```
Out[36]: trip_id                          int64
         starttime               datetime64[ns]
         stoptime                datetime64[ns]
         bikeid                          object
         tripduration                   float64
         from_station_name               object
         to_station_name                 object
         from_station_id                 object
         to_station_id                   object
         usertype                        object
         gender                          object
         birthyear                        int64
         Sex_num                        float64
         from_station_id_cat           category
         from_station_id_num               int8
         to_station_id_cat             category
         to_station_id_num                 int8
         Day                             object
         Day_cat                       category
         Day_num                           int8
         sthours                          int64
         stphours                         int64
         tripduration_minutes           float64
         age                              int32
         bmonth                           int64
         Date                            object
         dtype: object
```

```
In [37]: bikes.Day_num.tail()
```

```
Out[37]: 286842    6
         286843    6
         286844    6
         286845    6
         286847    6
         Name: Day_num, dtype: int8
```

8 Filtering Data for year 2015 and 2016

```
In [38]: # convert 'Time' to datetime format
         bikes['starttime'] = pd.to_datetime(bikes.starttime)
```

```
In [39]: # Breaking date column for further Analysis
         bikes['year'] = bikes['starttime'].dt.year
```

```
In [40]: # convert 'Time' to datetime format
         bikes['Date'] = pd.to_datetime(bikes.Date)
         bikes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 143543 entries, 0 to 286847
Data columns (total 27 columns):
trip_id                 143543 non-null int64
starttime               143543 non-null datetime64[ns]
stoptime                143543 non-null datetime64[ns]
bikeid                  143543 non-null object
tripduration            143543 non-null float64
from_station_name       143543 non-null object
to_station_name         143543 non-null object
from_station_id         143543 non-null object
to_station_id           143543 non-null object
usertype                143543 non-null object
gender                  143543 non-null object
birthyear               143543 non-null int64
Sex_num                 140905 non-null float64
from_station_id_cat     143543 non-null category
from_station_id_num     143543 non-null int8
to_station_id_cat       143543 non-null category
to_station_id_num       143543 non-null int8
Day                     143543 non-null object
Day_cat                 143543 non-null category
Day_num                 143543 non-null int8
sthours                 143543 non-null int64
stphours                143543 non-null int64
tripduration_minutes    143543 non-null float64
age                     143543 non-null int32
bmonth                  143543 non-null int64
Date                    143543 non-null datetime64[ns]
year                    143543 non-null int64
dtypes: category(3), datetime64[ns](3), float64(3), int32(1), int64(6), int8(3), object(8)
memory usage: 24.4+ MB
```

```
In [41]: # df2 belong to bikes dataset
         df2 = bikes[(bikes.year != 2014) ]
         df2.head(100)
```

```
Out[41]:        trip_id           starttime            stoptime    bikeid  \
         20240    25092 2015-01-01 00:37:00 2015-01-01 00:44:00  SEA00267
         20252    25114 2015-01-01 03:07:00 2015-01-01 03:18:00  SEA00472
         20255    25127 2015-01-01 08:12:00 2015-01-01 08:20:00  SEA00245
         20260    25132 2015-01-01 10:55:00 2015-01-01 11:03:00  SEA00391
         20263    25135 2015-01-01 11:31:00 2015-01-01 11:55:00  SEA00079
```

```
20265    25137 2015-01-01 11:45:00 2015-01-01 11:54:00  SEA00401
20270    25143 2015-01-01 12:13:00 2015-01-01 12:22:00  SEA00347
20271    25144 2015-01-01 12:14:00 2015-01-01 12:22:00  SEA00326
20274    25147 2015-01-01 12:16:00 2015-01-01 12:27:00  SEA00140
20276    25150 2015-01-01 12:52:00 2015-01-01 13:16:00  SEA00212
20277    25151 2015-01-01 12:52:00 2015-01-01 12:59:00  SEA00390
20281    25155 2015-01-01 12:58:00 2015-01-01 13:16:00  SEA00210
20282    25156 2015-01-01 12:59:00 2015-01-01 13:16:00  SEA00141
20292    25168 2015-01-01 13:17:00 2015-01-01 13:32:00  SEA00255
20325    25203 2015-01-01 14:22:00 2015-01-01 14:31:00  SEA00390
20327    25205 2015-01-01 14:27:00 2015-01-01 14:46:00  SEA00212
20330    25210 2015-01-01 14:49:00 2015-01-01 14:57:00  SEA00319
20331    25211 2015-01-01 15:00:00 2015-01-01 15:15:00  SEA00264
20332    25212 2015-01-01 15:02:00 2015-01-01 15:12:00  SEA00206
20333    25213 2015-01-01 15:15:00 2015-01-01 15:25:00  SEA00418
20337    25217 2015-01-01 15:19:00 2015-01-01 15:27:00  SEA00056
20342    25222 2015-01-01 15:27:00 2015-01-01 15:36:00  SEA00312
20346    25226 2015-01-01 15:40:00 2015-01-01 15:57:00  SEA00215
20347    25228 2015-01-01 15:46:00 2015-01-01 15:54:00  SEA00056
20348    25229 2015-01-01 15:45:00 2015-01-01 15:58:00  SEA00382
20350    25231 2015-01-01 16:04:00 2015-01-01 16:12:00  SEA00212
20351    25232 2015-01-01 16:04:00 2015-01-01 16:13:00  SEA00135
20356    25237 2015-01-01 16:17:00 2015-01-01 16:33:00  SEA00421
20357    25238 2015-01-01 16:19:00 2015-01-01 16:27:00  SEA00278
20358    25239 2015-01-01 16:23:00 2015-01-01 16:28:00  SEA00116
...        ...                 ...                 ...       ...
20437    25324 2015-01-02 10:42:00 2015-01-02 11:04:00  SEA00299
20438    25325 2015-01-02 11:21:00 2015-01-02 11:34:00  SEA00100
20439    25326 2015-01-02 11:25:00 2015-01-02 11:32:00  SEA00309
20440    25327 2015-01-02 11:38:00 2015-01-02 11:55:00  SEA00168
20443    25330 2015-01-02 11:52:00 2015-01-02 11:58:00  SEA00208
20446    25333 2015-01-02 12:13:00 2015-01-02 12:20:00  SEA00405
20447    25334 2015-01-02 12:21:00 2015-01-02 12:32:00  SEA00277
20448    25335 2015-01-02 12:26:00 2015-01-02 12:49:00  SEA00153
20449    25336 2015-01-02 12:26:00 2015-01-02 12:43:00  SEA00496
20450    25337 2015-01-02 12:39:00 2015-01-02 12:44:00  SEA00340
20451    25338 2015-01-02 12:43:00 2015-01-02 12:49:00  SEA00198
20453    25340 2015-01-02 12:50:00 2015-01-02 13:01:00  SEA00196
20454    25341 2015-01-02 12:52:00 2015-01-02 13:09:00  SEA00496
20455    25342 2015-01-02 12:52:00 2015-01-02 12:59:00  SEA00056
20457    25344 2015-01-02 12:56:00 2015-01-02 13:03:00  SEA00309
20458    25345 2015-01-02 13:02:00 2015-01-02 13:15:00  SEA00100
20459    25346 2015-01-02 13:08:00 2015-01-02 13:28:00  SEA00447
20461    25349 2015-01-02 13:16:00 2015-01-02 13:26:00  SEA00135
20464    25353 2015-01-02 13:26:00 2015-01-02 13:31:00  SEA00483
20465    25354 2015-01-02 13:38:00 2015-01-02 13:55:00  SEA00063
20467    25356 2015-01-02 13:56:00 2015-01-02 14:03:00  SEA00232
20468    25357 2015-01-02 14:03:00 2015-01-02 14:09:00  SEA00356
```

```
20469    25358 2015-01-02 14:10:00 2015-01-02 14:16:00  SEA00278
20470    25359 2015-01-02 14:12:00 2015-01-02 14:33:00  SEA00477
20474    25363 2015-01-02 14:32:00 2015-01-02 14:49:00  SEA00168
20476    25365 2015-01-02 14:48:00 2015-01-02 14:57:00  SEA00235
20477    25366 2015-01-02 14:53:00 2015-01-02 15:07:00  SEA00477
20479    25371 2015-01-02 15:39:00 2015-01-02 15:52:00  SEA00383
20481    25373 2015-01-02 15:51:00 2015-01-02 15:56:00  SEA00208
20489    25381 2015-01-02 16:31:00 2015-01-02 16:43:00  SEA00365

       tripduration                                  from_station_name  \
20240       459.469                             Harvard Ave & E Pine St
20252       614.453                               9th Ave N & Mercer St
20255       504.420                                E Pine St & 16th Ave
20260       470.801                                E Pine St & 16th Ave
20263      1461.638                              12th Ave & E Denny Way
20265       524.885                             Harvard Ave & E Pine St
20270       514.110                       E Harrison St & Broadway Ave E
20271       472.876                       E Harrison St & Broadway Ave E
20274       684.994                                   2nd Ave & Vine St
20276      1463.476                                E Pine St & 16th Ave
20277       397.101             Cal Anderson Park / 11th Ave & Pine St
20281      1101.863                        E Blaine St & Fairview Ave E
20282      1057.551                        E Blaine St & Fairview Ave E
20292       929.421         Lake Union Park / Valley St & Boren Ave N
20325       515.097                       E Harrison St & Broadway Ave E
20327      1149.250                               6th Ave S & S King St
20330       485.702                        E Blaine St & Fairview Ave E
20331       910.564                       E Harrison St & Broadway Ave E
20332       628.489                            Bellevue Ave & E Pine St
20333       549.486                       Summit Ave E & E Republican St
20337       465.429                                  3rd Ave & Broad St
20342       495.314                       E Harrison St & Broadway Ave E
20346       983.226                            15th Ave E & E Thomas St
20347       474.824                                   2nd Ave & Pine St
20348       737.699                            15th Ave E & E Thomas St
20350       480.341                                E Pine St & 16th Ave
20351       491.257                                E Pine St & 16th Ave
20356       939.485                       City Hall / 4th Ave & James St
20357       451.038                              12th Ave & E Denny Way
20358       311.966                         PATH / 9th Ave & Westlake Ave
...             ...                                                 ...
20437      1265.463  Occidental Park / Occidental Ave S & S Washing...
20438       835.595                             Harvard Ave & E Pine St
20439       406.067  Fred Hutchinson Cancer Research Center / Fairv...
20440       993.430                         Eastlake Ave E & E Allison St
20443       330.189                            12th Ave & NE Campus Pkwy
20446       416.075                                   2nd Ave & Vine St
20447       652.883  King Street Station Plaza / 2nd Ave Extension ...
```

```
20448      1372.678  Burke-Gilman Trail / NE Blakeley St & 24th Ave NE
20449      1012.147               Key Arena / 1st Ave N & Harrison St
20450       342.515                                2nd Ave & Pine St
20451       373.090                               3rd Ave & Broad St
20453       639.008  King Street Station Plaza / 2nd Ave Extension ...
20454      1048.554                               7th Ave & Union St
20455       403.213                               3rd Ave & Broad St
20457       376.005                     PATH / 9th Ave & Westlake Ave
20458       795.976                         15th Ave E & E Thomas St
20459      1150.649                      E Blaine St & Fairview Ave E
20461       649.109                     E Harrison St & Broadway Ave E
20464       350.746                                2nd Ave & Vine St
20465       993.728                     E Harrison St & Broadway Ave E
20467       425.978                              2nd Ave & Spring St
20468       330.987                           Westlake Ave & 6th Ave
20469       365.864                     PATH / 9th Ave & Westlake Ave
20470      1269.127  Occidental Park / Occidental Ave S & S Washing...
20474      1023.856            Lake Union Park / Valley St & Boren Ave N
20476       563.844  UW Engineering Library / E Stevens Way NE & Je...
20477       836.574               Cal Anderson Park / 11th Ave & Pine St
20479       785.410                                2nd Ave & Pine St
20481       318.126  UW Engineering Library / E Stevens Way NE & Je...
20489       752.358                           2nd Ave & Blanchard St

                                  to_station_name from_station_id  \
20240          Cal Anderson Park / 11th Ave & Pine St           CH-09
20252                    E Blaine St & Fairview Ave E          DPD-01
20255                               2nd Ave & Pine St           CH-07
20260                              7th Ave & Union St           CH-07
20263             Key Arena / 1st Ave N & Harrison St           CH-06
20265                  Summit Ave E & E Republican St           CH-09
20270      Seattle University / E Columbia St & 12th Ave           CH-02
20271      Seattle University / E Columbia St & 12th Ave           CH-02
20274                  Republican St & Westlake Ave N           BT-03
20276                          6th Ave S & S King St           CH-07
20277                  E Harrison St & Broadway Ave E           CH-08
20281                          Westlake Ave & 6th Ave           EL-03
20282                          Westlake Ave & 6th Ave           EL-03
20292             Key Arena / 1st Ave N & Harrison St          SLU-17
20325      Seattle University / E Columbia St & 12th Ave           CH-02
20327                           E Pine St & 16th Ave           ID-04
20330          Lake Union Park / Valley St & Boren Ave N           EL-03
20331                       12th Ave & NE Campus Pkwy           CH-02
20332                        15th Ave E & E Thomas St           CH-12
20333                        15th Ave E & E Thomas St           CH-03
20337                               2nd Ave & Pine St           BT-01
20342          Cal Anderson Park / 11th Ave & Pine St           CH-02
20346                    Eastlake Ave E & E Allison St           CH-05
```

```
20347                                 3rd Ave & Broad St         CBD-13
20348                          REI / Yale Ave N & John St          CH-05
20350                      E Harrison St & Broadway Ave E          CH-07
20351                      E Harrison St & Broadway Ave E          CH-07
20356                      E Harrison St & Broadway Ave E         CBD-07
20357                       PATH / 9th Ave & Westlake Ave          CH-06
20358                          REI / Yale Ave N & John St         SLU-07
...                                                   ...            ...
20437              Cal Anderson Park / 11th Ave & Pine St          PS-04
20438                           15th Ave E & E Thomas St          CH-09
20439                       PATH / 9th Ave & Westlake Ave          EL-01
20440          Lake Union Park / Valley St & Boren Ave N          EL-05
20443  UW Engineering Library / E Stevens Way NE & Je...          UD-04
20446                      Republican St & Westlake Ave N          BT-03
20447  Seattle Aquarium / Alaskan Way S & Elliott Bay...          PS-05
20448                           15th Ave NE & NE 40th St          UD-01
20449                                 7th Ave & Union St         SLU-19
20450                                1st Ave & Marion St         CBD-13
20451                       PATH / 9th Ave & Westlake Ave          BT-01
20453                                2nd Ave & Spring St          PS-05
20454                  Key Arena / 1st Ave N & Harrison St         CBD-03
20455                                2nd Ave & Spring St          BT-01
20457  Fred Hutchinson Cancer Research Center / Fairv...         SLU-07
20458                           Bellevue Ave & E Pine St          CH-05
20459                                2nd Ave & Spring St          EL-03
20461                                 E Pine St & 16th Ave          CH-02
20464                      Republican St & Westlake Ave N          BT-03
20465                                  2nd Ave & Pine St          CH-02
20467  King Street Station Plaza / 2nd Ave Extension ...         CBD-06
20468                                 3rd Ave & Broad St         SLU-15
20469                                 7th Ave & Union St         SLU-07
20470              Cal Anderson Park / 11th Ave & Pine St          PS-04
20474                       Eastlake Ave E & E Allison St         SLU-17
20476                          12th Ave & NE Campus Pkwy          UW-06
20477                                1st Ave & Marion St          CH-08
20479                        E Blaine St & Fairview Ave E         CBD-13
20481                      NE 42nd St & University Way NE          UW-06
20489                           Summit Ave & E Denny Way          BT-05

      to_station_id usertype  ...        Day   Day_cat  Day_num sthours  \
20240          CH-08   Member  ...   Thursday  Thursday        4       0
20252          EL-03   Member  ...   Thursday  Thursday        4       3
20255         CBD-13   Member  ...   Thursday  Thursday        4       8
20260         CBD-03   Member  ...   Thursday  Thursday        4      10
20263         SLU-19   Member  ...   Thursday  Thursday        4      11
20265          CH-03   Member  ...   Thursday  Thursday        4      11
20270          FH-04   Member  ...   Thursday  Thursday        4      12
20271          FH-04   Member  ...   Thursday  Thursday        4      12
```

```
20274          SLU-04   Member  ...   Thursday  Thursday        4       12
20276           ID-04   Member  ...   Thursday  Thursday        4       12
20277           CH-02   Member  ...   Thursday  Thursday        4       12
20281          SLU-15   Member  ...   Thursday  Thursday        4       12
20282          SLU-15   Member  ...   Thursday  Thursday        4       12
20292          SLU-19   Member  ...   Thursday  Thursday        4       13
20325           FH-04   Member  ...   Thursday  Thursday        4       14
20327           CH-07   Member  ...   Thursday  Thursday        4       14
20330          SLU-17   Member  ...   Thursday  Thursday        4       14
20331           UD-04   Member  ...   Thursday  Thursday        4       15
20332           CH-05   Member  ...   Thursday  Thursday        4       15
20333           CH-05   Member  ...   Thursday  Thursday        4       15
20337          CBD-13   Member  ...   Thursday  Thursday        4       15
20342           CH-08   Member  ...   Thursday  Thursday        4       15
20346           EL-05   Member  ...   Thursday  Thursday        4       15
20347           BT-01   Member  ...   Thursday  Thursday        4       15
20348          SLU-01   Member  ...   Thursday  Thursday        4       15
20350           CH-02   Member  ...   Thursday  Thursday        4       16
20351           CH-02   Member  ...   Thursday  Thursday        4       16
20356           CH-02   Member  ...   Thursday  Thursday        4       16
20357          SLU-07   Member  ...   Thursday  Thursday        4       16
20358          SLU-01   Member  ...   Thursday  Thursday        4       16
...               ...      ...  ...        ...       ...      ...      ...
20437           CH-08   Member  ...     Friday    Friday        0       10
20438           CH-05   Member  ...     Friday    Friday        0       11
20439          SLU-07   Member  ...     Friday    Friday        0       11
20440          SLU-17   Member  ...     Friday    Friday        0       11
20443           UW-06   Member  ...     Friday    Friday        0       11
20446          SLU-04   Member  ...     Friday    Friday        0       12
20447           WF-04   Member  ...     Friday    Friday        0       12
20448           UW-04   Member  ...     Friday    Friday        0       12
20449          CBD-03   Member  ...     Friday    Friday        0       12
20450          CBD-05   Member  ...     Friday    Friday        0       12
20451          SLU-07   Member  ...     Friday    Friday        0       12
20453          CBD-06   Member  ...     Friday    Friday        0       12
20454          SLU-19   Member  ...     Friday    Friday        0       12
20455          CBD-06   Member  ...     Friday    Friday        0       12
20457           EL-01   Member  ...     Friday    Friday        0       12
20458           CH-12   Member  ...     Friday    Friday        0       13
20459          CBD-06   Member  ...     Friday    Friday        0       13
20461           CH-07   Member  ...     Friday    Friday        0       13
20464          SLU-04   Member  ...     Friday    Friday        0       13
20465          CBD-13   Member  ...     Friday    Friday        0       13
20467           PS-05   Member  ...     Friday    Friday        0       13
20468           BT-01   Member  ...     Friday    Friday        0       14
20469          CBD-03   Member  ...     Friday    Friday        0       14
20470           CH-08   Member  ...     Friday    Friday        0       14
20474           EL-05   Member  ...     Friday    Friday        0       14
```

```
20476         UD-04   Member  ...     Friday    Friday         0       14
20477        CBD-05   Member  ...     Friday    Friday         0       14
20479         EL-03   Member  ...     Friday    Friday         0       15
20481         UD-02   Member  ...     Friday    Friday         0       15
20489         CH-01   Member  ...     Friday    Friday         0       16

       stphours tripduration_minutes  age bmonth       Date  year
20240         0                 7.66   27      1 2015-01-01  2015
20252         3                10.24   38      1 2015-01-01  2015
20255         8                 8.41   32      1 2015-01-01  2015
20260        11                 7.85   38      1 2015-01-01  2015
20263        11                24.36   31      1 2015-01-01  2015
20265        11                 8.75   31      1 2015-01-01  2015
20270        12                 8.57   32      1 2015-01-01  2015
20271        12                 7.88   37      1 2015-01-01  2015
20274        12                11.42   37      1 2015-01-01  2015
20276        13                24.39   33      1 2015-01-01  2015
20277        12                 6.62   43      1 2015-01-01  2015
20281        13                18.36   41      1 2015-01-01  2015
20282        13                17.63   44      1 2015-01-01  2015
20292        13                15.49   63      1 2015-01-01  2015
20325        14                 8.58   37      1 2015-01-01  2015
20327        14                19.15   33      1 2015-01-01  2015
20330        14                 8.10   35      1 2015-01-01  2015
20331        15                15.18   50      1 2015-01-01  2015
20332        15                10.47   32      1 2015-01-01  2015
20333        15                 9.16   41      1 2015-01-01  2015
20337        15                 7.76   49      1 2015-01-01  2015
20342        15                 8.26   43      1 2015-01-01  2015
20346        15                16.39   34      1 2015-01-01  2015
20347        15                 7.91   49      1 2015-01-01  2015
20348        15                12.29   30      1 2015-01-01  2015
20350        16                 8.01   37      1 2015-01-01  2015
20351        16                 8.19   32      1 2015-01-01  2015
20356        16                15.66   31      1 2015-01-01  2015
20357        16                 7.52   27      1 2015-01-01  2015
20358        16                 5.20   43      1 2015-01-01  2015
...         ...                  ...  ...    ...        ...   ...
20437        11                21.09   32      1 2015-01-02  2015
20438        11                13.93   71      1 2015-01-02  2015
20439        11                 6.77   55      1 2015-01-02  2015
20440        11                16.56   34      1 2015-01-02  2015
20443        11                 5.50   28      1 2015-01-02  2015
20446        12                 6.93   37      1 2015-01-02  2015
20447        12                10.88   35      1 2015-01-02  2015
20448        12                22.88   41      1 2015-01-02  2015
20449        12                16.87   47      1 2015-01-02  2015
20450        12                 5.71   69      1 2015-01-02  2015
```

```
20451         12                  6.22   34       1 2015-01-02  2015
20453         13                 10.65   35       1 2015-01-02  2015
20454         13                 17.48   47       1 2015-01-02  2015
20455         12                  6.72   34       1 2015-01-02  2015
20457         13                  6.27   55       1 2015-01-02  2015
20458         13                 13.27   71       1 2015-01-02  2015
20459         13                 19.18   41       1 2015-01-02  2015
20461         13                 10.82   50       1 2015-01-02  2015
20464         13                  5.85   29       1 2015-01-02  2015
20465         13                 16.56   27       1 2015-01-02  2015
20467         14                  7.10   35       1 2015-01-02  2015
20468         14                  5.52   34       1 2015-01-02  2015
20469         14                  6.10   45       1 2015-01-02  2015
20470         14                 21.15   69       1 2015-01-02  2015
20474         14                 17.06   34       1 2015-01-02  2015
20476         14                  9.40   31       1 2015-01-02  2015
20477         15                 13.94   69       1 2015-01-02  2015
20479         15                 13.09   41       1 2015-01-02  2015
20481         15                  5.30   28       1 2015-01-02  2015
20489         16                 12.54   36       1 2015-01-02  2015

[100 rows x 27 columns]
```

9 Saving Bike data

```
In [42]: df2.to_csv('C:\\Users\\mrferozi\\Desktop\\GitHub\\Bike\\dataset\\cycle\\trip_clean.cs
```

---

10 Cleaning Weather Data

```
In [5]: # read in a CSV
        df = pd.read_csv('C:/Users/mrferozi/Desktop/GitHub/Bike/cycle-share-dataset-orignal/we

In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 689 entries, 0 to 688
Data columns (total 21 columns):
Date                          689 non-null object
Max_Temperature_F             689 non-null int64
Mean_Temperature_F            688 non-null float64
Min_TemperatureF              689 non-null int64
Max_Dew_Point_F               689 non-null int64
MeanDew_Point_F               689 non-null int64
Min_Dewpoint_F                689 non-null int64
Max_Humidity                  689 non-null int64
```

```
Mean_Humidity                 689 non-null int64
Min_Humidity                  689 non-null int64
Max_Sea_Level_Pressure_In     689 non-null float64
Mean_Sea_Level_Pressure_In    689 non-null float64
Min_Sea_Level_Pressure_In     689 non-null float64
Max_Visibility_Miles          689 non-null int64
Mean_Visibility_Miles         689 non-null int64
Min_Visibility_Miles          689 non-null int64
Max_Wind_Speed_MPH            689 non-null int64
Mean_Wind_Speed_MPH           689 non-null int64
Max_Gust_Speed_MPH            504 non-null object
Precipitation_In              689 non-null float64
Events                        328 non-null object
dtypes: float64(5), int64(13), object(3)
memory usage: 113.1+ KB
```

11 Converting temprature from Fahrenheit to Celsius

The definition of precipitation is any form of water - liquid or solid - falling from the sky. It includes rain, sleet, snow, hail and drizzle plus a few less common occurrences such as ice pellets, diamond dust and freezing rain. https://www.metoffice.gov.uk/learning/learn-about-the-weather/weather-phenomena/what-is-precipitation

- Converting temprature from Fahrenheit to Celsius

  T(řC) = (T(řF) - 32) Œ 5/9
  Example
  Convert 68 degrees Fahrenheit to degrees Celsius:
  T(řC) = (68řF - 32) Œ 5/9 = 20 řC https://www.rapidtables.com/convert/temperature/how-fahrenheit-to-celsius.html

```
In [7]: a = df.Mean_Temperature_F-32
```

```
In [8]: df['Mean_Temperature_C'] = a * 5/9
        desred_decimals = 2
        df['Mean_Temperature_C'] = df['Mean_Temperature_C'].apply(lambda x: round(x,desred_dec
```

```
In [9]: # 'isnull' returns a DataFrame of booleans (True if missing, False if not missing)
        df.isnull().tail()
```

```
Out[9]:      Date Max_Temperature_F Mean_Temperature_F Min_TemperatureF  \
        684  False             False              False            False
        685  False             False              False            False
        686  False             False              False            False
        687  False             False              False            False
        688  False             False              False            False
```

```
    Max_Dew_Point_F MeanDew_Point_F Min_Dewpoint_F Max_Humidity Mean_Humidity  \
684           False           False          False        False         False
685           False           False          False        False         False
686           False           False          False        False         False
687           False           False          False        False         False
688           False           False          False        False         False

    Min_Humidity        ...         Min_Sea_Level_Pressure_In  \
684        False        ...                             False
685        False        ...                             False
686        False        ...                             False
687        False        ...                             False
688        False        ...                             False

    Max_Visibility_Miles Mean_Visibility_Miles Min_Visibility_Miles  \
684                False                 False                False
685                False                 False                False
686                False                 False                False
687                False                 False                False
688                False                 False                False

    Max_Wind_Speed_MPH Mean_Wind_Speed_MPH Max_Gust_Speed_MPH  \
684              False               False               True
685              False               False               True
686              False               False               True
687              False               False              False
688              False               False               True

    Precipitation_In Events Mean_Temperature_C
684            False   True              False
685            False   True              False
686            False   True              False
687            False   True              False
688            False  False              False

[5 rows x 22 columns]
```

```
In [10]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[10]: Date                          0
         Max_Temperature_F             0
         Mean_Temperature_F            1
         Min_TemperatureF              0
         Max_Dew_Point_F               0
         MeanDew_Point_F               0
         Min_Dewpoint_F                0
         Max_Humidity                  0
```

```
Mean_Humidity                   0
Min_Humidity                    0
Max_Sea_Level_Pressure_In       0
Mean_Sea_Level_Pressure_In      0
Min_Sea_Level_Pressure_In       0
Max_Visibility_Miles            0
Mean_Visibility_Miles           0
Min_Visibility_Miles            0
Max_Wind_Speed_MPH              0
Mean_Wind_Speed_MPH             0
Max_Gust_Speed_MPH            185
Precipitation_In                0
Events                        361
Mean_Temperature_C              1
dtype: int64
```

12 Filling Max_Gust_Speed_MPH with Some Values

```
In [11]: #Checking column for exsisting values
         df['Max_Gust_Speed_MPH']
```

```
Out[11]: 0        21
         1        17
         2        25
         3         -
         4         -
         5         -
         6        18
         7         -
         8        21
         9        22
         10       22
         11        -
         12       41
         13       30
         14        -
         15        -
         16       24
         17        -
         18        -
         19        -
         20        -
         21       20
         22        -
         23        -
         24       37
         25        -
```

```
26         -
27        16
28         -
29        22
      ...
659       17
660      NaN
661       21
662      NaN
663      NaN
664      NaN
665      NaN
666      NaN
667       17
668       20
669      NaN
670      NaN
671      NaN
672      NaN
673      NaN
674      NaN
675      NaN
676      NaN
677      NaN
678      NaN
679      NaN
680      NaN
681      NaN
682      NaN
683      NaN
684      NaN
685      NaN
686      NaN
687       18
688      NaN
Name: Max_Gust_Speed_MPH, dtype: object
```

```
In [12]: # fill in missing values with a specified value
         df['Max_Gust_Speed_MPH'].fillna(value='0', inplace=True)
         df.Max_Gust_Speed_MPH.replace('-','0',inplace=True)
         df['Max_Gust_Speed_MPH']
```

```
Out[12]: 0       21
         1       17
         2       25
         3        0
         4        0
         5        0
```

```
6       18
7        0
8       21
9       22
10      22
11       0
12      41
13      30
14       0
15       0
16      24
17       0
18       0
19       0
20       0
21      20
22       0
23       0
24      37
25       0
26       0
27      16
28       0
29      22
        ..
659     17
660      0
661     21
662      0
663      0
664      0
665      0
666      0
667     17
668     20
669      0
670      0
671      0
672      0
673      0
674      0
675      0
676      0
677      0
678      0
679      0
680      0
681      0
```

```
682      0
683      0
684      0
685      0
686      0
687     18
688      0
Name: Max_Gust_Speed_MPH, dtype: object
```

```
In [13]: ##Filling Null values with the mean of Max_Gust_Speed_MPH' values
         df['Max_Gust_Speed_MPH'] = df['Max_Gust_Speed_MPH'].astype(int)
         df['Max_Gust_Speed_MPH'].mean()
```

```
Out[13]: 9.252539912917271
```

```
In [14]: df.Max_Gust_Speed_MPH.replace('0','9.26',inplace=True)
```

13 Filling Events with Some Values

```
In [15]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[15]: Date                           0
         Max_Temperature_F              0
         Mean_Temperature_F             1
         Min_TemperatureF               0
         Max_Dew_Point_F                0
         MeanDew_Point_F                0
         Min_Dewpoint_F                 0
         Max_Humidity                   0
         Mean_Humidity                  0
         Min_Humidity                   0
         Max_Sea_Level_Pressure_In      0
         Mean_Sea_Level_Pressure_In     0
         Min_Sea_Level_Pressure_In      0
         Max_Visibility_Miles           0
         Mean_Visibility_Miles          0
         Min_Visibility_Miles           0
         Max_Wind_Speed_MPH             0
         Mean_Wind_Speed_MPH            0
         Max_Gust_Speed_MPH             0
         Precipitation_In               0
         Events                       361
         Mean_Temperature_C             1
         dtype: int64
```

```
In [16]: import pandas as pd
         Inplace = True
```

```
        for index, row in df.iterrows():
            a = row.Precipitation_In
            if a == 0:
                df.at[index, 'Events'] = 'Rain'
```

```
In [17]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[17]: Date                              0
         Max_Temperature_F                 0
         Mean_Temperature_F                1
         Min_TemperatureF                  0
         Max_Dew_Point_F                   0
         MeanDew_Point_F                   0
         Min_Dewpoint_F                    0
         Max_Humidity                      0
         Mean_Humidity                     0
         Min_Humidity                      0
         Max_Sea_Level_Pressure_In         0
         Mean_Sea_Level_Pressure_In        0
         Min_Sea_Level_Pressure_In         0
         Max_Visibility_Miles              0
         Mean_Visibility_Miles             0
         Min_Visibility_Miles              0
         Max_Wind_Speed_MPH                0
         Mean_Wind_Speed_MPH               0
         Max_Gust_Speed_MPH                0
         Precipitation_In                  0
         Events                           14
         Mean_Temperature_C                1
         dtype: int64
```

```
In [18]: import pandas as pd
         Inplace = True
         #df['Mean_Temperature_C'] = df['Mean_Temperature_C'].astype(int)
         for index, row in df.iterrows():
             a = row.Mean_Temperature_C
             if a >= 14:
                 df.at[index, 'Events'] = 'Sunny'
             print (a)
```

```
16.67
15.0
14.44
16.11
15.56
17.78
17.78
15.56
```

```
14.44
14.44
12.78
13.33
14.44
11.11
12.22
13.33
15.0
14.44
12.22
11.11
10.56
14.44
14.44
13.89
15.56
11.11
8.89
11.11
8.89
5.56
3.89
4.44
3.33
3.33
3.89
5.0
3.89
6.67
9.44
11.11
8.89
10.0
8.33
13.33
15.0
14.44
8.89
1.67
0.56
1.11
1.67
4.44
6.67
10.0
11.11
9.44
```

```
12.78
13.89
16.11
12.78
11.11
7.78
8.33
8.89
8.89
8.89
8.89
10.0
11.11
12.78
8.89
8.89
6.67
5.56
5.56
8.33
6.67
5.56
1.11
1.11
1.67
3.33
3.33
8.89
13.33
10.0
8.33
6.67
6.67
7.78
8.89
9.44
6.67
5.56
5.0
10.0
7.78
12.22
9.44
7.78
4.44
8.89
12.22
14.44
```

```
13.89
12.78
11.11
10.0
8.89
6.67
5.56
8.89
10.0
8.89
8.33
12.22
13.33
12.22
12.22
11.11
11.11
10.56
16.67
12.22
12.22
8.89
11.67
11.11
8.89
10.56
11.11
10.0
7.22
6.67
7.22
9.44
10.56
10.56
8.33
6.67
9.44
5.56
6.67
8.89
9.44
10.0
10.0
9.44
10.0
13.33
15.56
15.0
```

```
14.44
10.0
11.11
10.56
12.22
13.33
13.33
12.78
9.44
8.89
11.11
12.22
16.67
15.0
14.44
13.89
15.56
10.0
10.0
10.0
9.44
8.89
10.0
11.11
12.22
12.22
12.22
12.22
10.0
11.11
8.89
8.33
9.44
11.11
13.33
13.89
14.44
15.56
13.33
11.11
10.0
11.11
11.11
11.11
18.89
13.33
13.33
13.89
```

```
15.0
14.44
13.89
12.78
12.22
13.33
14.44
16.67
18.89
16.67
13.33
14.44
12.22
15.56
15.56
14.44
16.67
18.89
18.33
17.78
20.0
15.56
15.0
15.56
14.44
17.78
19.44
21.11
21.67
17.22
18.33
15.56
15.56
16.67
18.33
20.0
22.22
23.33
22.78
22.22
20.0
18.89
17.78
17.78
20.56
22.22
17.78
19.44
```

```
21.11
18.89
18.89
21.11
19.44
20.0
22.22
24.44
24.44
25.56
24.44
25.56
23.89
24.44
26.67
26.67
25.0
25.56
23.33
21.11
23.33
22.78
20.0
20.0
22.22
22.22
22.22
22.22
22.22
21.67
25.56
28.33
23.33
20.0
20.0
21.67
19.44
18.89
18.89
18.33
22.22
23.89
25.56
26.67
25.0
25.0
24.44
21.11
```

```
18.33
21.11
22.22
21.11
22.22
23.89
24.44
23.89
23.33
18.33
18.89
20.0
21.11
23.33
25.56
20.0
18.89
20.0
21.11
18.89
20.0
21.67
23.33
21.11
18.89
18.33
18.89
17.78
17.22
15.56
15.56
14.44
14.44
18.89
19.44
20.56
20.0
22.22
21.67
18.89
14.44
15.0
15.56
16.67
17.78
18.89
18.89
15.56
```

```
13.89
15.56
17.78
16.11
15.56
12.78
16.11
16.67
15.56
15.56
14.44
16.67
16.67
16.67
15.56
15.56
17.78
16.67
18.89
15.56
16.11
14.44
14.44
15.56
15.56
16.67
15.56
16.67
15.0
13.89
12.78
11.11
12.78
13.89
12.78
12.22
13.33
15.56
15.56
15.0
12.22
10.0
8.33
7.22
11.11
13.33
12.22
11.11
```

```
8.89
7.78
10.0
8.89
13.33
8.89
6.67
6.67
12.22
7.22
6.67
5.56
6.67
5.56
4.44
6.67
5.56
5.0
3.89
2.22
1.11
2.22
8.33
8.89
11.11
9.44
8.33
11.11
11.11
14.44
11.11
10.0
8.89
8.89
7.78
5.56
4.44
5.0
5.56
7.78
6.67
7.78
5.56
5.56
5.0
5.0
3.33
3.33
```

```
4.44
4.44
4.44
2.22
1.67
1.11
1.11
1.11
3.33
4.44
6.11
3.33
3.89
3.89
6.11
3.33
5.56
8.89
5.56
4.44
8.89
7.78
8.89
6.67
7.78
8.89
10.0
7.22
6.67
7.22
10.0
12.22
10.0
8.33
5.56
5.56
5.0
5.56
6.67
7.78
8.89
5.56
6.11
8.89
10.0
10.0
10.0
11.11
```

```
8.89
nan
12.22
11.67
11.11
9.44
7.78
6.67
5.56
7.78
8.89
10.0
10.0
10.56
11.11
8.89
8.89
10.0
10.0
11.11
8.89
13.33
11.11
8.89
7.78
8.89
8.89
7.78
7.78
8.33
6.67
7.78
8.33
8.33
11.11
12.22
11.11
10.0
10.0
10.0
9.44
10.0
10.0
8.33
7.78
10.0
12.22
15.0
```

```
14.44
12.22
14.44
12.22
11.11
13.33
17.22
17.78
12.78
11.11
12.22
11.11
11.11
11.11
12.22
13.33
17.78
21.11
20.0
19.44
17.22
14.44
14.44
10.56
11.11
11.11
13.33
14.44
12.78
13.33
17.22
21.11
18.33
13.33
15.56
18.89
20.0
13.33
13.89
15.56
18.89
17.78
20.0
15.0
13.33
13.33
16.11
15.56
```

```
12.78
14.44
13.33
15.56
15.0
16.67
15.56
14.44
13.89
13.33
14.44
15.56
18.89
18.89
17.78
20.0
22.22
25.56
23.33
22.22
16.67
15.0
14.44
14.44
16.67
15.56
12.22
13.33
16.11
15.56
15.56
14.44
18.89
17.78
18.33
16.67
15.56
17.78
19.44
22.78
18.89
18.33
18.33
18.89
20.56
17.78
16.67
17.22
```

```
18.89
17.78
18.89
17.78
18.89
17.78
20.0
20.0
18.89
18.89
18.89
21.11
19.44
21.11
22.22
21.67
18.89
18.89
20.0
22.78
21.11
21.67
22.78
23.89
19.44
17.78
18.89
17.78
18.89
20.56
21.11
17.22
18.89
17.78
18.89
20.0
21.11
23.89
25.56
21.67
22.22
20.56
21.11
22.78
15.56
25.56
18.89
17.78
```

```
19.44
21.11
25.0
23.33
18.89
20.0
20.0
17.78
18.33
```

```
In [1]: #Search google for Sunny condition in Seattle
        from IPython.display import HTML
        HTML('<iframe src=https://www.accuweather.com/en/us/seattle-wa/98104/current-weather/35

Out[1]: <IPython.core.display.HTML object>

In [20]: # count the number of missing values in each Series
         df.isnull().sum()

Out[20]: Date                            0
         Max_Temperature_F               0
         Mean_Temperature_F              1
         Min_TemperatureF                0
         Max_Dew_Point_F                 0
         MeanDew_Point_F                 0
         Min_Dewpoint_F                  0
         Max_Humidity                    0
         Mean_Humidity                   0
         Min_Humidity                    0
         Max_Sea_Level_Pressure_In       0
         Mean_Sea_Level_Pressure_In      0
         Min_Sea_Level_Pressure_In       0
         Max_Visibility_Miles            0
         Mean_Visibility_Miles           0
         Min_Visibility_Miles            0
         Max_Wind_Speed_MPH              0
         Mean_Wind_Speed_MPH             0
         Max_Gust_Speed_MPH              0
         Precipitation_In                0
         Events                          7
         Mean_Temperature_C              1
         dtype: int64

In [21]: # confirm that the missing values were filled in
         df['Events'].value_counts().head()

Out[21]: Rain                 339
         Sunny                325
```

```
Fog , Rain             5
Fog-Rain               5
Rain-Thunderstorm      3
Name: Events, dtype: int64
```

14 Encoding Columns values

```
In [22]: df.Events.replace('Fog-Rain', 'Fog, Rain',inplace=True)
         df.Events.replace('Fog , Rain', 'Fog, Rain',inplace=True)
```

```
In [23]: df['Events'].value_counts().head()
```

```
Out[23]: Rain                 339
         Sunny                325
         Fog, Rain             10
         Rain-Thunderstorm      3
         Fog                    2
         Name: Events, dtype: int64
```

15 Converting Categorical variable to nominal

```
In [24]: df["Events_cat"] = df["Events"].astype('category')
         df.dtypes
```

```
Out[24]: Date                            object
         Max_Temperature_F                int64
         Mean_Temperature_F             float64
         Min_TemperatureF                 int64
         Max_Dew_Point_F                  int64
         MeanDew_Point_F                  int64
         Min_Dewpoint_F                   int64
         Max_Humidity                     int64
         Mean_Humidity                    int64
         Min_Humidity                     int64
         Max_Sea_Level_Pressure_In      float64
         Mean_Sea_Level_Pressure_In     float64
         Min_Sea_Level_Pressure_In      float64
         Max_Visibility_Miles             int64
         Mean_Visibility_Miles            int64
         Min_Visibility_Miles             int64
         Max_Wind_Speed_MPH               int64
         Mean_Wind_Speed_MPH              int64
         Max_Gust_Speed_MPH              object
         Precipitation_In               float64
         Events                          object
         Mean_Temperature_C             float64
         Events_cat                    category
         dtype: object
```

```
In [25]: df["Events_num"] = df["Events_cat"].cat.codes

In [29]: df["Events_cat"] = df["Events"].astype('category')
         df["Events_num"] = df["Events_cat"].cat.codes

In [36]: df['Events_num'].mode()

Out[36]: 0    2
         dtype: int8

In [37]: df[(df.Events_num == 2)]

Out[37]:            Date  Max_Temperature_F  Mean_Temperature_F  Min_TemperatureF  \
         10   10/23/2014                 62                55.0                50
         11   10/24/2014                 60                56.0                51
         13   10/26/2014                 59                52.0                48
         14   10/27/2014                 62                54.0                45
         15   10/28/2014                 62                56.0                51
         18   10/31/2014                 59                54.0                52
         19   11/01/2014                 55                52.0                48
         20   11/02/2014                 57                51.0                45
         23   11/05/2014                 61                57.0                53
         25   11/07/2014                 59                52.0                45
         26   11/08/2014                 55                48.0                42
         27   11/09/2014                 57                52.0                48
         28   11/10/2014                 54                48.0                43
         29   11/11/2014                 48                42.0                36
         30   11/12/2014                 46                39.0                32
         31   11/13/2014                 48                40.0                30
         32   11/14/2014                 48                38.0                28
         33   11/15/2014                 48                38.0                27
         34   11/16/2014                 50                39.0                28
         35   11/17/2014                 52                41.0                30
         36   11/18/2014                 48                39.0                30
         37   11/19/2014                 52                44.0                36
         38   11/20/2014                 54                49.0                44
         39   11/21/2014                 55                52.0                50
         40   11/22/2014                 51                48.0                45
         41   11/23/2014                 57                50.0                44
         42   11/24/2014                 55                47.0                39
         43   11/25/2014                 60                56.0                51
         46   11/28/2014                 57                48.0                39
         48   11/30/2014                 39                33.0                27
         ..          ...                ...                 ...               ...
         533   3/29/2016                 64                50.0                37
         534   3/30/2016                 68                54.0                43
         537  04/02/2016                 63                54.0                48
         539  04/04/2016                 60                54.0                46
         540  04/05/2016                 57                52.0                46
```

```
541  04/06/2016                 68                 56.0                 45
544  04/09/2016                 66                 55.0                 46
545  04/10/2016                 59                 52.0                 48
546  04/11/2016                 57                 54.0                 50
547  04/12/2016                 60                 52.0                 46
548   4/13/2016                 59                 52.0                 46
549   4/14/2016                 59                 52.0                 45
550   4/15/2016                 60                 54.0                 48
551   4/16/2016                 68                 56.0                 43
559   4/24/2016                 55                 51.0                 48
561   4/26/2016                 60                 52.0                 43
562   4/27/2016                 61                 56.0                 51
564   4/29/2016                 61                 55.0                 50
565   4/30/2016                 68                 56.0                 45
569  05/04/2016                 60                 56.0                 54
573  05/08/2016                 63                 56.0                 50
574  05/09/2016                 66                 57.0                 48
580   5/15/2016                 57                 56.0                 54
581   5/16/2016                 60                 56.0                 52
584   5/19/2016                 63                 55.0                 48
586   5/21/2016                 59                 56.0                 52
592   5/27/2016                 64                 57.0                 51
593   5/28/2016                 64                 56.0                 50
610   6/14/2016                 60                 54.0                 48
611   6/15/2016                 66                 56.0                 45

     Max_Dew_Point_F  MeanDew_Point_F  Min_Dewpoint_F  Max_Humidity  \
10                49               47              44            86
11                50               47              44            86
13                48               44              42            86
14                46               43              41            87
15                54               50              45            88
18                52               49              46            89
19                48               45              42            89
20                52               47              42            90
23                55               51              48            90
25                45               43              40            86
26                44               42              39            93
27                50               46              43            87
28                43               40              34            86
29                39               18               3            80
30                10                4               1            40
31                19                9               2            64
32                26               18              11            69
33                25               18              10            75
34                27               21              15            69
35                29               25              22            75
36                36               31              25            86
```

```
37              37           34           30          86
38              46           43           39          87
39              48           45           41          87
40              43           41           39          86
41              47           42           39          82
42              50           41           35          89
43              54           50           46          89
46              51           43           35          87
48              23           21           18          78
..             ...          ...          ...         ...
533             43           39           36         100
534             45           42           39          93
537             49           47           45         100
539             50           43           40          92
540             46           44           42          86
541             47           44           40          90
544             50           47           44         100
545             45           43           42          86
546             44           42           39          80
547             48           44           41          97
548             46           42           37          97
549             46           42           38          93
550             47           43           39          90
551             45           41           35          89
559             49           46           42          93
561             44           41           37          93
562             47           44           40          80
564             49           46           43          96
565             47           44           40          97
569             50           47           44          80
573             50           46           42         100
574             47           44           38          90
580             55           53           51          97
581             53           50           49         100
584             50           48           43          97
586             51           38           28          90
592             49           44           37          83
593             56           48           31          96
610             47           45           42          86
611             47           42           38          90

    Mean_Humidity  Min_Humidity    ...      Mean_Visibility_Miles  \
10             76            62    ...                         10
11             75            60    ...                         10
13             71            62    ...                         10
14             72            48    ...                         10
15             80            72    ...                          8
18             83            67    ...                          8
```

```
19          78         62    ...              10
20          82         62    ...               9
23          79         64    ...              10
25          72         53    ...              10
26          78         60    ...               8
27          80         72    ...               9
28          71         54    ...              10
29          42         20    ...              10
30          24         16    ...              10
31          30         18    ...              10
32          48         23    ...              10
33          51         21    ...              10
34          53         25    ...              10
35          60         34    ...              10
36          75         54    ...               8
37          66         48    ...               9
38          80         71    ...              10
39          79         66    ...               8
40          77         66    ...              10
41          74         57    ...              10
42          79         66    ...               8
43          83         77    ...               6
46          80         72    ...               7
48          64         47    ...              10
..         ...        ...    ...             ...
533         74         38    ...               9
534         64         37    ...              10
537         80         58    ...              10
539         72         51    ...              10
540         77         64    ...              10
541         67         38    ...              10
544         73         48    ...              10
545         71         57    ...              10
546         67         51    ...              10
547         79         59    ...               9
548         73         44    ...              10
549         73         49    ...              10
550         70         48    ...              10
551         62         31    ...              10
559         83         66    ...               9
561         69         44    ...              10
562         67         46    ...              10
564         74         53    ...              10
565         64         37    ...              10
569         72         58    ...              10
573         72         46    ...               9
574         64         37    ...              10
580         91         86    ...              10
```

```
581              82            69      ...                        8
584              83            52      ...                        9
586              56            32      ...                        9
592              67            38      ...                       10
593              78            37      ...                       10
610              73            53      ...                       10
611              63            36      ...                       10

     Min_Visibility_Miles  Max_Wind_Speed_MPH  Mean_Wind_Speed_MPH  \
10                     10                  15                    9
11                      8                   8                    4
13                     10                  20                   12
14                     10                   7                    4
15                      2                  14                    5
18                      2                   8                    4
19                      7                   6                    1
20                      4                   9                    5
23                      3                   7                    2
25                     10                   7                    3
26                      2                   7                    1
27                      3                  12                    6
28                     10                   8                    2
29                     10                  15                    5
30                     10                  13                    8
31                     10                  12                    4
32                     10                   8                    2
33                     10                   7                    1
34                     10                   8                    1
35                      9                   8                    1
36                      5                   6                    1
37                      5                   9                    2
38                      7                   8                    4
39                      4                  18                    7
40                     10                  13                    8
41                      5                  20                   10
42                      2                  13                    6
43                      2                  17                    6
46                      2                  15                    9
48                     10                   8                    2
..                    ...                 ...                  ...
533                     1                  12                    4
534                    10                   8                    2
537                    10                  12                    4
539                     5                  18                   12
540                    10                  14                    8
541                    10                  12                    4
544                    10                   6                    2
545                    10                   9                    7
```

```
546                 10              10                10
547                  4              16                 8
548                 10              13                 6
549                  6              14                 6
550                 10              14                 6
551                 10               8                 4
559                  5              10                 6
561                 10               8                 4
562                 10              12                 2
564                  5               8                 4
565                 10              12                 3
569                 10              14                 5
573                  2              13                 7
574                 10               8                 5
580                  6               8                 6
581                  3               6                 3
584                  4              13                 7
586                  4              10                 3
592                 10              12                 7
593                  4              14                 9
610                 10              21                10
611                 10              10                 7

     Max_Gust_Speed_MPH  Precipitation_In  Events  Mean_Temperature_C  \
10                   22              0.35    Rain               12.78
11                 9.26              0.13    Rain               13.33
13                   30              0.05    Rain               11.11
14                 9.26              0.01    Rain               12.22
15                 9.26              0.34    Rain               13.33
18                 9.26              0.77    Rain               12.22
19                 9.26              0.00    Rain               11.11
20                 9.26              0.11    Rain               10.56
23                 9.26              0.27    Rain               13.89
25                 9.26              0.00    Rain               11.11
26                 9.26              0.00    Rain                8.89
27                   16              0.29    Rain               11.11
28                 9.26              0.00    Rain                8.89
29                   22              0.00    Rain                5.56
30                   22              0.00    Rain                3.89
31                   22              0.00    Rain                4.44
32                 9.26              0.00    Rain                3.33
33                 9.26              0.00    Rain                3.33
34                 9.26              0.00    Rain                3.89
35                 9.26              0.00    Rain                5.00
36                 9.26              0.00    Rain                3.89
37                 9.26              0.00    Rain                6.67
38                 9.26              0.11    Rain                9.44
39                   33              0.67    Rain               11.11
```

```
40                  21             0.03    Rain            8.89
41                  28             0.42    Rain           10.00
42                  25             0.01    Rain            8.33
43                  29             0.33    Rain           13.33
46                  26             1.39    Rain            8.89
48                9.26             0.00    Rain            0.56
..                 ...              ...     ...             ...
533                 20             0.00    Rain           10.00
534               9.26             0.00    Rain           12.22
537               9.26             0.00    Rain           12.22
539                 30             0.17    Rain           12.22
540                 21             0.00    Rain           11.11
541                 17             0.00    Rain           13.33
544               9.26             0.00    Rain           12.78
545               9.26             0.00    Rain           11.11
546               9.26             0.00    Rain           12.22
547                 23             0.44    Rain           11.11
548                 17             0.00    Rain           11.11
549                 32             0.20    Rain           11.11
550                 17             0.00    Rain           12.22
551               9.26             0.00    Rain           13.33
559                 18             0.27    Rain           10.56
561               9.26             0.00    Rain           11.11
562               9.26             0.00    Rain           13.33
564               9.26             0.04    Rain           12.78
565                 18             0.00    Rain           13.33
569               9.26             0.00    Rain           13.33
573                 16             0.02    Rain           13.33
574               9.26             0.00    Rain           13.89
580               9.26             0.00    Rain           13.33
581               9.26             0.00    Rain           13.33
584                 20             0.22    Rain           12.78
586               9.26             0.04    Rain           13.33
592               9.26             0.02    Rain           13.89
593                 23             0.03    Rain           13.33
610                 28             0.05    Rain           12.22
611                 16             0.00    Rain           13.33

    Events_cat  Events_num
10        Rain           2
11        Rain           2
13        Rain           2
14        Rain           2
15        Rain           2
18        Rain           2
19        Rain           2
20        Rain           2
23        Rain           2
```

```
25         Rain           2
26         Rain           2
27         Rain           2
28         Rain           2
29         Rain           2
30         Rain           2
31         Rain           2
32         Rain           2
33         Rain           2
34         Rain           2
35         Rain           2
36         Rain           2
37         Rain           2
38         Rain           2
39         Rain           2
40         Rain           2
41         Rain           2
42         Rain           2
43         Rain           2
46         Rain           2
48         Rain           2
..          ...         ...
533        Rain           2
534        Rain           2
537        Rain           2
539        Rain           2
540        Rain           2
541        Rain           2
544        Rain           2
545        Rain           2
546        Rain           2
547        Rain           2
548        Rain           2
549        Rain           2
550        Rain           2
551        Rain           2
559        Rain           2
561        Rain           2
562        Rain           2
564        Rain           2
565        Rain           2
569        Rain           2
573        Rain           2
574        Rain           2
580        Rain           2
581        Rain           2
584        Rain           2
586        Rain           2
```

```
592       Rain           2
593       Rain           2
610       Rain           2
611       Rain           2

[339 rows x 24 columns]
```

As we can see that rain is our mean Event We will fill all empty event with 6

```
In [38]: # fill in missing values with a specified value
         df['Events'].fillna(value='Rain', inplace=True)
```

The Events_num is appera as -1 which is not right, to fix this problem we repeate above step

```
In [39]: df["Events_cat"] = df["Events"].astype('category')
         df["Events_num"] = df["Events_cat"].cat.codes
```

```
In [40]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[40]: Date                          0
         Max_Temperature_F             0
         Mean_Temperature_F            1
         Min_TemperatureF              0
         Max_Dew_Point_F               0
         MeanDew_Point_F               0
         Min_Dewpoint_F                0
         Max_Humidity                  0
         Mean_Humidity                 0
         Min_Humidity                  0
         Max_Sea_Level_Pressure_In     0
         Mean_Sea_Level_Pressure_In    0
         Min_Sea_Level_Pressure_In     0
         Max_Visibility_Miles          0
         Mean_Visibility_Miles         0
         Min_Visibility_Miles          0
         Max_Wind_Speed_MPH            0
         Mean_Wind_Speed_MPH           0
         Max_Gust_Speed_MPH            0
         Precipitation_In              0
         Events                        0
         Mean_Temperature_C            1
         Events_cat                    0
         Events_num                    0
         dtype: int64
```

```
In [41]: # confirm that the missing values were filled in
         df['Events'].value_counts().head()
```

```
Out[41]: Rain                   346
         Sunny                  325
         Fog, Rain               10
         Rain-Thunderstorm        3
         Fog                      2
         Name: Events, dtype: int64
```

```
In [42]: # if 'any' values are missing in a row, then drop that row
         df.dropna(how='any').shape
```

```
Out[42]: (688, 24)
```

```
In [43]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[43]: Date                             0
         Max_Temperature_F                0
         Mean_Temperature_F               1
         Min_TemperatureF                 0
         Max_Dew_Point_F                  0
         MeanDew_Point_F                  0
         Min_Dewpoint_F                   0
         Max_Humidity                     0
         Mean_Humidity                    0
         Min_Humidity                     0
         Max_Sea_Level_Pressure_In        0
         Mean_Sea_Level_Pressure_In       0
         Min_Sea_Level_Pressure_In        0
         Max_Visibility_Miles             0
         Mean_Visibility_Miles            0
         Min_Visibility_Miles             0
         Max_Wind_Speed_MPH               0
         Mean_Wind_Speed_MPH              0
         Max_Gust_Speed_MPH               0
         Precipitation_In                 0
         Events                           0
         Mean_Temperature_C               1
         Events_cat                       0
         Events_num                       0
         dtype: int64
```

```
In [44]: # fill in missing values with a specified value
         df['Mean_Temperature_F'].fillna(value='57', inplace=True)
```

```
In [45]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[45]: Date                             0
         Max_Temperature_F                0
         Mean_Temperature_F            0
         Min_TemperatureF              0
         Max_Dew_Point_F               0
         MeanDew_Point_F               0
         Min_Dewpoint_F                0
         Max_Humidity                  0
         Mean_Humidity                 0
         Min_Humidity                  0
         Max_Sea_Level_Pressure_In     0
         Mean_Sea_Level_Pressure_In    0
         Min_Sea_Level_Pressure_In     0
         Max_Visibility_Miles          0
         Mean_Visibility_Miles         0
         Min_Visibility_Miles          0
         Max_Wind_Speed_MPH            0
         Mean_Wind_Speed_MPH           0
         Max_Gust_Speed_MPH            0
         Precipitation_In              0
         Events                        0
         Mean_Temperature_C            1
         Events_cat                    0
         Events_num                    0
         dtype: int64
```

```
In [46]: df['Mean_Temperature_C'].mean()
```

```
Out[46]: 13.657965116279073
```

```
In [47]: df['Mean_Temperature_C'].fillna(value='14', inplace=True)
```

```
In [48]: # count the number of missing values in each Series
         df.isnull().sum()
```

```
Out[48]: Date                          0
         Max_Temperature_F             0
         Mean_Temperature_F            0
         Min_TemperatureF              0
         Max_Dew_Point_F               0
         MeanDew_Point_F               0
         Min_Dewpoint_F                0
         Max_Humidity                  0
         Mean_Humidity                 0
         Min_Humidity                  0
         Max_Sea_Level_Pressure_In     0
         Mean_Sea_Level_Pressure_In    0
         Min_Sea_Level_Pressure_In     0
         Max_Visibility_Miles          0
         Mean_Visibility_Miles         0
         Min_Visibility_Miles          0
```

```
Max_Wind_Speed_MPH          0
Mean_Wind_Speed_MPH         0
Max_Gust_Speed_MPH          0
Precipitation_In            0
Events                      0
Mean_Temperature_C          0
Events_cat                  0
Events_num                  0
dtype: int64
```

```
In [49]: df['Mean_Temperature_F'] = df['Mean_Temperature_F'].astype(int)
         df['Mean_Temperature_F'].mean()
```

```
Out[49]: 56.58490566037736
```

```
In [50]: # 'isnull' returns a DataFrame of booleans (True if missing, False if not missing)
         df.isnull().tail()
```

```
Out[50]:       Date Max_Temperature_F Mean_Temperature_F Min_TemperatureF  \
         684  False             False              False            False
         685  False             False              False            False
         686  False             False              False            False
         687  False             False              False            False
         688  False             False              False            False

             Max_Dew_Point_F MeanDew_Point_F Min_Dewpoint_F Max_Humidity Mean_Humidity  \
         684           False           False          False        False         False
         685           False           False          False        False         False
         686           False           False          False        False         False
         687           False           False          False        False         False
         688           False           False          False        False         False

             Min_Humidity    ...     Mean_Visibility_Miles Min_Visibility_Miles  \
         684        False    ...                     False                False
         685        False    ...                     False                False
         686        False    ...                     False                False
         687        False    ...                     False                False
         688        False    ...                     False                False

             Max_Wind_Speed_MPH Mean_Wind_Speed_MPH Max_Gust_Speed_MPH  \
         684              False               False              False
         685              False               False              False
         686              False               False              False
         687              False               False              False
         688              False               False              False

             Precipitation_In Events Mean_Temperature_C Events_cat Events_num
         684            False  False              False      False      False
         685            False  False              False      False      False
```

```
686              False  False               False      False      False
687              False  False               False      False      False
688              False  False               False      False      False

[5 rows x 24 columns]
```

16 Encoding Data

```
In [51]: # convert 'Time' to datetime format
         df['Date'] = pd.to_datetime(df.Date)
```

```
In [52]: df['month'] = df['Date'].dt.month
```

```
In [53]: # remove a single column (axis=1 refers to columns)
         df.drop('Events_cat', axis=1, inplace=True)
```

```
In [54]: df['year']=df['Date'].dt.year
```

17 Removing Duplicated Tuples

```
In [55]: # detect duplicate Date codes: True if an item is identical to a previous item
         df.Date.duplicated().tail()
```

```
Out[55]: 684    False
         685    False
         686    False
         687    False
         688    False
         Name: Date, dtype: bool
```

```
In [56]: # count the duplicate items (True becomes 1, False becomes 0)
         df.Date.duplicated().sum()
```

```
Out[56]: 0
```

The above codes shows that we do not have any duplicated rows in our dataset

```
In [57]: df.head(100)
```

```
Out[57]:          Date  Max_Temperature_F  Mean_Temperature_F  Min_TemperatureF  \
         0  2014-10-13                 71                  62                54
         1  2014-10-14                 63                  59                55
         2  2014-10-15                 62                  58                54
         3  2014-10-16                 71                  61                52
         4  2014-10-17                 64                  60                57
         5  2014-10-18                 68                  64                59
         6  2014-10-19                 73                  64                55
         7  2014-10-20                 66                  60                55
```

```
8   2014-10-21            64            58            55
9   2014-10-22            60            58            57
10  2014-10-23            62            55            50
11  2014-10-24            60            56            51
12  2014-10-25            64            58            52
13  2014-10-26            59            52            48
14  2014-10-27            62            54            45
15  2014-10-28            62            56            51
16  2014-10-29            64            59            54
17  2014-10-30            62            58            55
18  2014-10-31            59            54            52
19  2014-11-01            55            52            48
20  2014-11-02            57            51            45
21  2014-11-03            60            58            55
22  2014-11-04            61            58            55
23  2014-11-05            61            57            53
24  2014-11-06            64            60            54
25  2014-11-07            59            52            45
26  2014-11-08            55            48            42
27  2014-11-09            57            52            48
28  2014-11-10            54            48            43
29  2014-11-11            48            42            36
..         ...           ...           ...           ...
70  2014-12-22            54            48            42
71  2014-12-23            53            48            43
72  2014-12-24            46            44            41
73  2014-12-25            48            42            37
74  2014-12-26            44            42            39
75  2014-12-27            52            47            42
76  2014-12-28            46            44            41
77  2014-12-29            46            42            35
78  2014-12-30            39            34            30
79  2014-12-31            41            34            26
80  2015-01-01            43            35            27
81  2015-01-02            44            38            32
82  2015-01-03            43            38            33
83  2015-01-04            54            48            41
84  2015-01-05            57            56            54
85  2015-01-06            55            50            46
86  2015-01-07            50            47            44
87  2015-01-08            48            44            39
88  2015-01-09            50            44            39
89  2015-01-10            48            46            44
90  2015-01-11            51            48            46
91  2015-01-12            53            49            45
92  2015-01-13            48            44            41
93  2015-01-14            46            42            37
94  2015-01-15            46            41            36
```

```
95 2015-01-16                55                    50                     44
96 2015-01-17                55                    46                     39
97 2015-01-18                60                    54                     48
98 2015-01-19                52                    49                     46
99 2015-01-20                51                    46                     39

    Max_Dew_Point_F  MeanDew_Point_F  Min_Dewpoint_F  Max_Humidity  \
0                55               51              46            87
1                52               51              50            88
2                53               50              46            87
3                49               46              42            83
4                55               51              41            87
5                59               57              55            90
6                57               55              53            94
7                57               54              50            90
8                52               49              46            87
9                55               53              48            88
10               49               47              44            86
11               50               47              44            86
12               53               49              44            87
13               48               44              42            86
14               46               43              41            87
15               54               50              45            88
16               54               51              50            88
17               55               53              50            88
18               52               49              46            89
19               48               45              42            89
20               52               47              42            90
21               53               51              48            84
22               54               51              48            90
23               55               51              48            90
24               55               50              45            90
25               45               43              40            86
26               44               42              39            93
27               50               46              43            87
28               43               40              34            86
29               39               18               3            80
..              ...              ...             ...           ...
70               45               40              37            86
71               52               42              36            96
72               39               38              37            89
73               39               37              34            87
74               37               36              34            87
75               45               40              36            87
76               39               38              35            85
77               36               28              21            87
78               24               17              14            73
79               25               22              18            78
```

```
80                  28                  25                 21                81
81                  39                  32                 25                86
82                  38                  35                 30                93
83                  48                  41                 37                89
84                  52                  50                 48                88
85                  48                  46                 43                93
86                  46                  43                 41                93
87                  41                  38                 36                89
88                  41                  39                 37                93
89                  45                  44                 41                93
90                  46                  45                 42                89
91                  43                  41                 39                90
92                  41                  38                 37                87
93                  38                  34                 28                89
94                  41                  34                 27                86
95                  45                  40                 36                87
96                  52                  41                 35                90
97                  51                  44                 42                82
98                  43                  42                 40                83
99                  39                  38                 34                87

    Mean_Humidity  Min_Humidity  ...   Min_Visibility_Miles  \
0              68            46  ...                      4
1              78            63  ...                      3
2              77            67  ...                      3
3              61            36  ...                     10
4              72            46  ...                      6
5              83            68  ...                      2
6              74            52  ...                      6
7              78            67  ...                      5
8              70            58  ...                      6
9              81            67  ...                      2
10             76            62  ...                     10
11             75            60  ...                      8
12             78            58  ...                      6
13             71            62  ...                     10
14             72            48  ...                     10
15             80            72  ...                      2
16             76            60  ...                      9
17             82            75  ...                      2
18             83            67  ...                      2
19             78            62  ...                      7
20             82            62  ...                      4
21             80            75  ...                      3
22             79            67  ...                      3
23             79            64  ...                      3
24             75            58  ...                      4
25             72            53  ...                     10
```

```
26             78          60  ...                2
27             80          72  ...                3
28             71          54  ...               10
29             42          20  ...               10
..            ...         ...  ...              ...
70             72          55  ...               10
71             80          66  ...                2
72             80          71  ...                7
73             79          66  ...                6
74             80          71  ...                4
75             78          69  ...                5
76             79          71  ...                8
77             62          40  ...               10
78             51          35  ...               10
79             60          41  ...               10
80             70          49  ...                6
81             76          67  ...                7
82             84          76  ...                3
83             84          77  ...                3
84             81          74  ...                2
85             84          74  ...                0
86             85          74  ...                1
87             82          71  ...                0
88             86          68  ...                0
89             88          83  ...                1
90             85          80  ...                0
91             78          64  ...                4
92             80          66  ...                2
93             77          60  ...                1
94             73          63  ...                3
95             71          57  ...               10
96             84          73  ...                3
97             73          67  ...                3
98             76          66  ...                7
99             76          59  ...                9

    Max_Wind_Speed_MPH  Mean_Wind_Speed_MPH  Max_Gust_Speed_MPH  \
0                   13                    4                  21
1                   10                    5                  17
2                   18                    7                  25
3                    9                    4                9.26
4                    8                    3                9.26
5                   10                    4                9.26
6                   10                    3                  18
7                   12                    5                9.26
8                   15                    8                  21
9                   14                    8                  22
10                  15                    9                  22
```

11	8	4	9.26
12	24	6	41
13	20	12	30
14	7	4	9.26
15	14	5	9.26
16	15	7	24
17	8	2	9.26
18	8	4	9.26
19	6	1	9.26
20	9	5	9.26
21	13	8	20
22	10	6	9.26
23	7	2	9.26
24	20	8	37
25	7	3	9.26
26	7	1	9.26
27	12	6	16
28	8	2	9.26
29	15	5	22
..	...	...	...
70	8	1	9.26
71	14	5	9.26
72	8	5	9.26
73	7	3	9.26
74	8	4	9.26
75	15	8	25
76	8	5	9.26
77	9	3	9.26
78	6	1	9.26
79	7	2	9.26
80	4	0	9.26
81	12	4	9.26
82	7	2	9.26
83	21	7	32
84	20	14	31
85	6	2	9.26
86	6	2	9.26
87	8	3	9.26
88	6	2	9.26
89	7	1	9.26
90	5	1	9.26
91	10	1	9.26
92	10	2	9.26
93	10	3	9.26
94	16	3	9.26
95	13	8	22
96	17	1	23
97	22	14	36

```
98                 10                    5                 9.26
99                  9                    1                 9.26

   Precipitation_In     Events  Mean_Temperature_C  Events_num month  year
0              0.00      Sunny               16.67           6    10  2014
1              0.11      Sunny                  15           6    10  2014
2              0.45      Sunny               14.44           6    10  2014
3              0.00      Sunny               16.11           6    10  2014
4              0.14      Sunny               15.56           6    10  2014
5              0.31      Sunny               17.78           6    10  2014
6              0.00      Sunny               17.78           6    10  2014
7              0.44      Sunny               15.56           6    10  2014
8              0.10      Sunny               14.44           6    10  2014
9              1.43      Sunny               14.44           6    10  2014
10             0.35       Rain               12.78           2    10  2014
11             0.13       Rain               13.33           2    10  2014
12             0.37      Sunny               14.44           6    10  2014
13             0.05       Rain               11.11           2    10  2014
14             0.01       Rain               12.22           2    10  2014
15             0.34       Rain               13.33           2    10  2014
16             0.04      Sunny                  15           6    10  2014
17             0.67      Sunny               14.44           6    10  2014
18             0.77       Rain               12.22           2    10  2014
19             0.00       Rain               11.11           2    11  2014
20             0.11       Rain               10.56           2    11  2014
21             0.24      Sunny               14.44           6    11  2014
22             0.05      Sunny               14.44           6    11  2014
23             0.27       Rain               13.89           2    11  2014
24             0.22      Sunny               15.56           6    11  2014
25             0.00       Rain               11.11           2    11  2014
26             0.00       Rain                8.89           2    11  2014
27             0.29       Rain               11.11           2    11  2014
28             0.00       Rain                8.89           2    11  2014
29             0.00       Rain                5.56           2    11  2014
..              ...        ...                 ...         ...   ...   ...
70             0.00       Rain                8.89           2    12  2014
71             0.61       Rain                8.89           2    12  2014
72             0.12       Rain                6.67           2    12  2014
73             0.00       Rain                5.56           2    12  2014
74             0.00       Rain                5.56           2    12  2014
75             0.12       Rain                8.33           2    12  2014
76             0.06       Rain                6.67           2    12  2014
77             0.00       Rain                5.56           2    12  2014
78             0.00       Rain                1.11           2    12  2014
79             0.00       Rain                1.11           2    12  2014
80             0.00       Rain                1.67           2     1  2015
81             0.03       Rain                3.33           2     1  2015
82             0.00       Rain                3.33           2     1  2015
```

```
         83             0.22       Rain              8.89          2      1  2015
         84             0.07       Rain             13.33          2      1  2015
         85             0.01        Fog                10          0      1  2015
         86             0.00       Rain              8.33          2      1  2015
         87             0.00       Rain              6.67          2      1  2015
         88             0.01  Fog, Rain              6.67          1      1  2015
         89             0.18       Rain              7.78          2      1  2015
         90             0.06  Fog, Rain              8.89          1      1  2015
         91             0.00       Rain              9.44          2      1  2015
         92             0.00       Rain              6.67          2      1  2015
         93             0.00       Rain              5.56          2      1  2015
         94             0.43       Rain                 5          2      1  2015
         95             0.00       Rain                10          2      1  2015
         96             0.76       Rain              7.78          2      1  2015
         97             0.23       Rain             12.22          2      1  2015
         98             0.03       Rain              9.44          2      1  2015
         99             0.00       Rain              7.78          2      1  2015

         [100 rows x 25 columns]
```

```
In [58]: df['year'] = df['year'].astype(int)

In [59]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 689 entries, 0 to 688
Data columns (total 25 columns):
Date                          689 non-null datetime64[ns]
Max_Temperature_F             689 non-null int64
Mean_Temperature_F            689 non-null int32
Min_TemperatureF              689 non-null int64
Max_Dew_Point_F               689 non-null int64
MeanDew_Point_F               689 non-null int64
Min_Dewpoint_F                689 non-null int64
Max_Humidity                  689 non-null int64
Mean_Humidity                 689 non-null int64
Min_Humidity                  689 non-null int64
Max_Sea_Level_Pressure_In     689 non-null float64
Mean_Sea_Level_Pressure_In    689 non-null float64
Min_Sea_Level_Pressure_In     689 non-null float64
Max_Visibility_Miles          689 non-null int64
Mean_Visibility_Miles         689 non-null int64
Min_Visibility_Miles          689 non-null int64
Max_Wind_Speed_MPH            689 non-null int64
Mean_Wind_Speed_MPH           689 non-null int64
Max_Gust_Speed_MPH            689 non-null object
Precipitation_In              689 non-null float64
Events                        689 non-null object
```

```
Mean_Temperature_C             689 non-null object
Events_num                     689 non-null int8
month                          689 non-null int64
year                           689 non-null int32
dtypes: datetime64[ns](1), float64(4), int32(2), int64(14), int8(1), object(3)
memory usage: 124.6+ KB
```

```
In [60]: df.Date.tail()
```

```
Out[60]: 684   2016-08-27
         685   2016-08-28
         686   2016-08-29
         687   2016-08-30
         688   2016-08-31
         Name: Date, dtype: datetime64[ns]
```

```
In [61]: # df1 belong to weather dataset
         df1 = df[(df.year != 2014)]
         df1.tail()
```

```
Out[61]:           Date  Max_Temperature_F  Mean_Temperature_F  Min_TemperatureF  \
         684 2016-08-27                 72                  66                61
         685 2016-08-28                 75                  68                59
         686 2016-08-29                 81                  68                55
         687 2016-08-30                 70                  64                57
         688 2016-08-31                 71                  65                59

              Max_Dew_Point_F  MeanDew_Point_F  Min_Dewpoint_F  Max_Humidity  \
         684               57               54              50            81
         685               54               52              50            80
         686               55               53              50            89
         687               55               53              52            83
         688               61               56              52            90

              Mean_Humidity  Min_Humidity  ...   Min_Visibility_Miles  \
         684             65            46  ...                     10
         685             65            44  ...                     10
         686             65            39  ...                      6
         687             69            53  ...                     10
         688             77            63  ...                      8

              Max_Wind_Speed_MPH  Mean_Wind_Speed_MPH  Max_Gust_Speed_MPH  \
         684                  16                    9                9.26
         685                  12                    9                9.26
         686                   9                    4                9.26
         687                  14                    9                  18
         688                  14                    8                9.26
```

```
     Precipitation_In  Events  Mean_Temperature_C  Events_num month  year
684               0.0   Sunny               18.89           6     8  2016
685               0.0   Sunny                  20           6     8  2016
686               0.0   Sunny                  20           6     8  2016
687               0.0   Sunny               17.78           6     8  2016
688               0.0   Sunny               18.33           6     8  2016

[5 rows x 25 columns]
```

18 Saving Weather Data

Weather data consist on 600 observations where bikes data is consisting over a hundred thousand of tuples which makles it hard to combine both data set for further weather analysis. To achieve this we will, break both weather and bikes data Date columns and applying programming technique to achieve weather value for each bicycle trip.

```
In [62]: df1.isnull().sum()
```

```
Out[62]: Date                              0
         Max_Temperature_F                 0
         Mean_Temperature_F                0
         Min_TemperatureF                  0
         Max_Dew_Point_F                   0
         MeanDew_Point_F                   0
         Min_Dewpoint_F                    0
         Max_Humidity                      0
         Mean_Humidity                     0
         Min_Humidity                      0
         Max_Sea_Level_Pressure_In         0
         Mean_Sea_Level_Pressure_In        0
         Min_Sea_Level_Pressure_In         0
         Max_Visibility_Miles              0
         Mean_Visibility_Miles             0
         Min_Visibility_Miles              0
         Max_Wind_Speed_MPH                0
         Mean_Wind_Speed_MPH               0
         Max_Gust_Speed_MPH                0
         Precipitation_In                  0
         Events                            0
         Mean_Temperature_C                0
         Events_num                        0
         month                             0
         year                              0
         dtype: int64
```

19 Data type Conversion

```
In [63]: df1['Mean_Temperature_C'] = df1['Mean_Temperature_C'].astype(int)
         df1.dtypes
```

```
Out[63]: Date                              datetime64[ns]
         Max_Temperature_F                          int64
         Mean_Temperature_F                         int32
         Min_TemperatureF                           int64
         Max_Dew_Point_F                            int64
         MeanDew_Point_F                            int64
         Min_Dewpoint_F                             int64
         Max_Humidity                               int64
         Mean_Humidity                              int64
         Min_Humidity                               int64
         Max_Sea_Level_Pressure_In                float64
         Mean_Sea_Level_Pressure_In               float64
         Min_Sea_Level_Pressure_In                float64
         Max_Visibility_Miles                       int64
         Mean_Visibility_Miles                      int64
         Min_Visibility_Miles                       int64
         Max_Wind_Speed_MPH                         int64
         Mean_Wind_Speed_MPH                        int64
         Max_Gust_Speed_MPH                        object
         Precipitation_In                         float64
         Events                                    object
         Mean_Temperature_C                         int32
         Events_num                                  int8
         month                                      int64
         year                                       int32
         dtype: object
```

```
In [64]: df1
```

```
Out[64]:           Date  Max_Temperature_F  Mean_Temperature_F  Min_TemperatureF  \
         80  2015-01-01                 43                  35                27
         81  2015-01-02                 44                  38                32
         82  2015-01-03                 43                  38                33
         83  2015-01-04                 54                  48                41
         84  2015-01-05                 57                  56                54
         85  2015-01-06                 55                  50                46
         86  2015-01-07                 50                  47                44
         87  2015-01-08                 48                  44                39
         88  2015-01-09                 50                  44                39
         89  2015-01-10                 48                  46                44
         90  2015-01-11                 51                  48                46
         91  2015-01-12                 53                  49                45
         92  2015-01-13                 48                  44                41
         93  2015-01-14                 46                  42                37
```

```
94   2015-01-15                 46                  41                  36
95   2015-01-16                 55                  50                  44
96   2015-01-17                 55                  46                  39
97   2015-01-18                 60                  54                  48
98   2015-01-19                 52                  49                  46
99   2015-01-20                 51                  46                  39
100  2015-01-21                 48                  40                  33
101  2015-01-22                 52                  48                  45
102  2015-01-23                 57                  54                  50
103  2015-01-24                 62                  58                  55
104  2015-01-25                 64                  57                  50
105  2015-01-26                 64                  55                  46
106  2015-01-27                 55                  52                  50
107  2015-01-28                 55                  50                  46
108  2015-01-29                 55                  48                  42
109  2015-01-30                 48                  44                  39
..          ...                ...                 ...                 ...
659  2016-08-02                 73                  64                  57
660  2016-08-03                 75                  66                  60
661  2016-08-04                 80                  69                  59
662  2016-08-05                 79                  70                  61
663  2016-08-06                 72                  63                  55
664  2016-08-07                 72                  66                  60
665  2016-08-08                 73                  64                  57
666  2016-08-09                 72                  66                  60
667  2016-08-10                 75                  68                  60
668  2016-08-11                 80                  70                  61
669  2016-08-12                 87                  75                  63
670  2016-08-13                 90                  78                  66
671  2016-08-14                 82                  71                  60
672  2016-08-15                 84                  72                  61
673  2016-08-16                 81                  69                  57
674  2016-08-17                 79                  70                  60
675  2016-08-18                 86                  73                  60
676  2016-08-19                 95                  60                  26
677  2016-08-20                 91                  78                  64
678  2016-08-21                 73                  66                  60
679  2016-08-22                 72                  64                  57
680  2016-08-23                 79                  67                  55
681  2016-08-24                 82                  70                  57
682  2016-08-25                 93                  77                  61
683  2016-08-26                 88                  74                  60
684  2016-08-27                 72                  66                  61
685  2016-08-28                 75                  68                  59
686  2016-08-29                 81                  68                  55
687  2016-08-30                 70                  64                  57
688  2016-08-31                 71                  65                  59
```

```
    Max_Dew_Point_F  MeanDew_Point_F  Min_Dewpoint_F  Max_Humidity  \
80               28               25              21            81   
81               39               32              25            86   
82               38               35              30            93   
83               48               41              37            89   
84               52               50              48            88   
85               48               46              43            93   
86               46               43              41            93   
87               41               38              36            89   
88               41               39              37            93   
89               45               44              41            93   
90               46               45              42            89   
91               43               41              39            90   
92               41               38              37            87   
93               38               34              28            89   
94               41               34              27            86   
95               45               40              36            87   
96               52               41              35            90   
97               51               44              42            82   
98               43               42              40            83   
99               39               38              34            87   
100              39               35              29            93   
101              46               43              38            87   
102              52               49              45            89   
103              55               52              51            87   
104              53               49              45            86   
105              47               46              42            93   
106              49               47              45            88   
107              45               42              40            87   
108              40               38              36            85   
109              40               38              35            96   
..              ...              ...             ...           ...   
659              53               51              50            78   
660              53               52              51            72   
661              56               53              51            81   
662              54               51              46            75   
663              51               49              48            77   
664              56               52              49            83   
665              54               53              48            84   
666              55               54              53            83   
667              56               55              54            83   
668              57               56              54            81   
669              59               57              55            78   
670              60               56              51            78   
671              54               52              44            78   
672              57               55              53            78   
673              56               54              51            90   
674              54               53              50            80   
```

```
675                 63                59                49              72
676                 77                54                 9              88
677                 57                53                48              70
678                 55                54                52              83
679                 59                51                46              93
680                 55                51                48              83
681                 57                55                54              89
682                 59                54                50              81
683                 57                52                45              82
684                 57                54                50              81
685                 54                52                50              80
686                 55                53                50              89
687                 55                53                52              83
688                 61                56                52              90

     Mean_Humidity  Min_Humidity  ...  Min_Visibility_Miles  \
80              70            49  ...                     6
81              76            67  ...                     7
82              84            76  ...                     3
83              84            77  ...                     3
84              81            74  ...                     2
85              84            74  ...                     0
86              85            74  ...                     1
87              82            71  ...                     0
88              86            68  ...                     0
89              88            83  ...                     1
90              85            80  ...                     0
91              78            64  ...                     4
92              80            66  ...                     2
93              77            60  ...                     1
94              73            63  ...                     3
95              71            57  ...                    10
96              84            73  ...                     3
97              73            67  ...                     3
98              76            66  ...                     7
99              76            59  ...                     9
100             81            66  ...                     0
101             81            76  ...                     5
102             84            76  ...                     2
103             82            75  ...                    10
104             77            54  ...                     6
105             77            52  ...                     0
106             83            80  ...                     2
107             76            64  ...                     7
108             71            57  ...                    10
109             85            71  ...                     0
..             ...           ...  ...                   ...
659             66            49  ...                    10
```

```
660                 62              45  ...                        10
661                 58              38  ...                        10
662                 56              32  ...                        10
663                 62              45  ...                        10
664                 69              49  ...                        10
665                 69              44  ...                        10
666                 70              53  ...                        10
667                 67              48  ...                        10
668                 63              47  ...                        10
669                 56              36  ...                        10
670                 52              28  ...                        10
671                 51              25  ...                        10
672                 57              37  ...                        10
673                 62              35  ...                        10
674                 56              40  ...                        10
675                 57              45  ...                        10
676                 48              23  ...                        10
677                 44              26  ...                        10
678                 68              48  ...                        10
679                 64              40  ...                        10
680                 58              39  ...                        10
681                 60              39  ...                        10
682                 51              22  ...                        10
683                 45              22  ...                        10
684                 65              46  ...                        10
685                 65              44  ...                        10
686                 65              39  ...                         6
687                 69              53  ...                        10
688                 77              63  ...                         8

     Max_Wind_Speed_MPH  Mean_Wind_Speed_MPH  Max_Gust_Speed_MPH  \
80                    4                    0                9.26
81                   12                    4                9.26
82                    7                    2                9.26
83                   21                    7                  32
84                   20                   14                  31
85                    6                    2                9.26
86                    6                    2                9.26
87                    8                    3                9.26
88                    6                    2                9.26
89                    7                    1                9.26
90                    5                    1                9.26
91                   10                    1                9.26
92                   10                    2                9.26
93                   10                    3                9.26
94                   16                    3                9.26
95                   13                    8                  22
96                   17                    1                  23
```

```
97                    22                    14                    36
98                    10                     5                  9.26
99                     9                     1                  9.26
100                    5                     1                  9.26
101                    8                     1                  9.26
102                   13                     6                    21
103                   10                     7                    20
104                    8                     3                  9.26
105                   13                     4                  9.26
106                   10                     6                  9.26
107                    6                     2                  9.26
108                    8                     1                  9.26
109                    8                     1                  9.26
..                   ...                   ...                   ...
659                   12                     6                    17
660                   10                     6                  9.26
661                   13                     5                    21
662                    9                     4                  9.26
663                   10                     6                  9.26
664                    9                     4                  9.26
665                    9                     4                  9.26
666                   10                     3                  9.26
667                   12                     2                    17
668                   12                     4                    20
669                   10                     3                  9.26
670                    8                     2                  9.26
671                   10                     4                  9.26
672                    9                     2                  9.26
673                    8                     2                  9.26
674                   12                     4                  9.26
675                   16                     8                  9.26
676                   16                     8                  9.26
677                   10                     3                  9.26
678                   14                     8                  9.26
679                   14                     8                  9.26
680                   17                     7                  9.26
681                   20                     5                  9.26
682                   14                     4                  9.26
683                   10                     4                  9.26
684                   16                     9                  9.26
685                   12                     9                  9.26
686                    9                     4                  9.26
687                   14                     9                    18
688                   14                     8                  9.26

     Precipitation_In     Events  Mean_Temperature_C  Events_num month  year
80               0.00       Rain                   1           2     1  2015
81               0.03       Rain                   3           2     1  2015
```

```
82           0.00       Rain           3      2    1  2015
83           0.22       Rain           8      2    1  2015
84           0.07       Rain          13      2    1  2015
85           0.01        Fog          10      0    1  2015
86           0.00       Rain           8      2    1  2015
87           0.00       Rain           6      2    1  2015
88           0.01  Fog, Rain           6      1    1  2015
89           0.18       Rain           7      2    1  2015
90           0.06  Fog, Rain           8      1    1  2015
91           0.00       Rain           9      2    1  2015
92           0.00       Rain           6      2    1  2015
93           0.00       Rain           5      2    1  2015
94           0.43       Rain           5      2    1  2015
95           0.00       Rain          10      2    1  2015
96           0.76       Rain           7      2    1  2015
97           0.23       Rain          12      2    1  2015
98           0.03       Rain           9      2    1  2015
99           0.00       Rain           7      2    1  2015
100          0.00       Rain           4      2    1  2015
101          0.03       Rain           8      2    1  2015
102          0.08       Rain          12      2    1  2015
103          0.02      Sunny          14      6    1  2015
104          0.01       Rain          13      2    1  2015
105          0.00       Rain          12      2    1  2015
106          0.02       Rain          11      2    1  2015
107          0.00       Rain          10      2    1  2015
108          0.00       Rain           8      2    1  2015
109          0.00       Rain           6      2    1  2015
..            ...        ...         ...    ...  ...   ...
659          0.00      Sunny          17      6    8  2016
660          0.00      Sunny          18      6    8  2016
661          0.00      Sunny          20      6    8  2016
662          0.00      Sunny          21      6    8  2016
663          0.00      Sunny          17      6    8  2016
664          0.03      Sunny          18      6    8  2016
665          0.00      Sunny          17      6    8  2016
666          0.00      Sunny          18      6    8  2016
667          0.00      Sunny          20      6    8  2016
668          0.00      Sunny          21      6    8  2016
669          0.00      Sunny          23      6    8  2016
670          0.00      Sunny          25      6    8  2016
671          0.00      Sunny          21      6    8  2016
672          0.00      Sunny          22      6    8  2016
673          0.00      Sunny          20      6    8  2016
674          0.00      Sunny          21      6    8  2016
675          0.00      Sunny          22      6    8  2016
676          0.00      Sunny          15      6    8  2016
677          0.00      Sunny          25      6    8  2016
```

```
678            0.00     Sunny               18          6     8  2016
679            0.00     Sunny               17          6     8  2016
680            0.00     Sunny               19          6     8  2016
681            0.00     Sunny               21          6     8  2016
682            0.00     Sunny               25          6     8  2016
683            0.00     Sunny               23          6     8  2016
684            0.00     Sunny               18          6     8  2016
685            0.00     Sunny               20          6     8  2016
686            0.00     Sunny               20          6     8  2016
687            0.00     Sunny               17          6     8  2016
688            0.00     Sunny               18          6     8  2016

[609 rows x 25 columns]
```

```
In [65]: df1.to_csv('C:\\Users\\mrferozi\\Desktop\\GitHub\\Bike\\dataset\\cycle\\weather_clean
```

19.1 Resources

References:*From the video series: Introduction to machine learning with scikit-learn* - scikit-learn documentation: Cross-validation, Model evaluation - scikit-learn issue on GitHub: MSE is negative when returned by cross_val_score - Section 5.1 of An Introduction to Statistical Learning (11 pages) and related videos: K-fold and leave-one-out cross-validation (14 minutes), Cross-validation the right and wrong ways (10 minutes) - Scott Fortmann-Roe: Accurately Measuring Model Prediction Error - Machine Learning Mastery: An Introduction to Feature Selection - Harvard CS109: Cross-Validation: The Right and Wrong Way - Journal of Cheminformatics: Cross-validation pitfalls when selecting and assessing regression and classification models