Proyek Analisis Data: Nama dataset

- Nama: Muhammad Amien Ramdhani
- Email: dhanimacbull@gmail.com
- Id Dicoding: amienramdhani

Menentukan Pertanyaan Bisnis

- Berapa Kecepatan Rata rata angin pada tahun 2011
- Berapa jumlah penyewa sepeda yang telah terdaftar dan yang belum terdaftar di tiap tahun

Menyiapkan semua library yang dibuthkan

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Data Wrangling

Gathering Data

```
day_df = pd.read_csv('day.csv')
day_df.head()
```

```
   instant      dteday  season  yr  mnth  holiday  weekday  workingday
\
0        1  2011-01-01       1   0     1        0        6           0

1        2  2011-01-02       1   0     1        0        0           0

2        3  2011-01-03       1   0     1        0        1           1

3        4  2011-01-04       1   0     1        0        2           1

4        5  2011-01-05       1   0     1        0        3           1


   weathersit      temp     atemp       hum  windspeed  casual
registered  \
0           2  0.344167  0.363625  0.805833   0.160446     331
654
1           2  0.363478  0.353739  0.696087   0.248539     131
670
2           1  0.196364  0.189405  0.437273   0.248309     120
1229
3           1  0.200000  0.212122  0.590435   0.160296     108
```

```
1454
4           1  0.226957  0.229270  0.436957   0.186900      82
1518
```

```
    cnt
0   985
1   801
2  1349
3  1562
4  1600
```

Assessing Data

Menilai Tabel day_df

```
day_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   instant     731 non-null    int64  
 1   dteday      731 non-null    object 
 2   season      731 non-null    int64  
 3   yr          731 non-null    int64  
 4   mnth        731 non-null    int64  
 5   holiday     731 non-null    int64  
 6   weekday     731 non-null    int64  
 7   workingday  731 non-null    int64  
 8   weathersit  731 non-null    int64  
 9   temp        731 non-null    float64
 10  atemp       731 non-null    float64
 11  hum         731 non-null    float64
 12  windspeed   731 non-null    float64
 13  casual      731 non-null    int64  
 14  registered  731 non-null    int64  
 15  cnt         731 non-null    int64  
dtypes: float64(4), int64(11), object(1)
memory usage: 91.5+ KB
```

```
day_df.isna().sum()
```

```
instant       0
dteday        0
season        0
yr            0
mnth          0
holiday       0
weekday       0
```

```
workingday    0
weathersit    0
temp          0
atemp         0
hum           0
windspeed     0
casual        0
registered    0
cnt           0
dtype: int64
```

```
print("Jumlah duplikasi: ", day_df.duplicated().sum())
```

```
Jumlah duplikasi:  0
```

```
day_df.describe()
```

```
          instant      season          yr        mnth     holiday
weekday  \
count  731.000000  731.000000  731.000000  731.000000  731.000000
731.000000
mean   366.000000    2.496580    0.500684    6.519836    0.028728
2.997264
std    211.165812    1.110807    0.500342    3.451913    0.167155
2.004787
min      1.000000    1.000000    0.000000    1.000000    0.000000
0.000000
25%    183.500000    2.000000    0.000000    4.000000    0.000000
1.000000
50%    366.000000    3.000000    1.000000    7.000000    0.000000
3.000000
75%    548.500000    3.000000    1.000000   10.000000    0.000000
5.000000
max    731.000000    4.000000    1.000000   12.000000    1.000000
6.000000

       workingday  weathersit        temp       atemp         hum
windspeed  \
count  731.000000  731.000000  731.000000  731.000000  731.000000
731.000000
mean     0.683995    1.395349    0.495385    0.474354    0.627894
0.190486
std      0.465233    0.544894    0.183051    0.162961    0.142429
0.077498
min      0.000000    1.000000    0.059130    0.079070    0.000000
0.022392
25%      0.000000    1.000000    0.337083    0.337842    0.520000
0.134950
50%      1.000000    1.000000    0.498333    0.486733    0.626667
0.180975
```

```
75%      1.000000    2.000000    0.655417    0.608602    0.730209
0.233214
max      1.000000    3.000000    0.861667    0.840896    0.972500
0.507463

            casual   registered          cnt
count   731.000000   731.000000   731.000000
mean    848.176471  3656.172367  4504.348837
std     686.622488  1560.256377  1937.211452
min       2.000000    20.000000    22.000000
25%     315.500000  2497.000000  3152.000000
50%     713.000000  3662.000000  4548.000000
75%    1096.000000  4776.500000  5956.000000
max    3410.000000  6946.000000  8714.000000
```

Cleaning Data

Membersihkan day_df

Mengubah tipe data dteday menjadi tipe data datetime

```
day_df['dteday'] = pd.to_datetime(day_df['dteday'])

day_df['temp'] = day_df['temp']*41
day_df['atemp'] = day_df['atemp']*50
day_df['hum'] = day_df['hum']*100
day_df['windspeed'] = day_df['windspeed']*67

col = ['season', 'holiday', 'weekday',
       'workingday', 'weathersit']

for i in col:
  if i in day_df.columns.to_list():
    day_df[i] = day_df[i].astype('category')

day_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   instant     731 non-null    int64
 1   dteday      731 non-null    datetime64[ns]
 2   season      731 non-null    category
 3   yr          731 non-null    int64
 4   mnth        731 non-null    int64
 5   holiday     731 non-null    category
 6   weekday     731 non-null    category
 7   workingday  731 non-null    category
```

```
 8   weathersit  731 non-null    category
 9   temp        731 non-null    float64
 10  atemp       731 non-null    float64
 11  hum         731 non-null    float64
 12  windspeed   731 non-null    float64
 13  casual      731 non-null    int64
 14  registered  731 non-null    int64
 15  cnt         731 non-null    int64
dtypes: category(5), datetime64[ns](1), float64(4), int64(6)
memory usage: 67.4 KB
```

```
day_df.head(5)
```

```
   instant     dteday season  yr  mnth holiday weekday workingday
weathersit  \
0        1 2011-01-01      1   0     1       0       6          0
2
1        2 2011-01-02      1   0     1       0       0          0
2
2        3 2011-01-03      1   0     1       0       1          1
1
3        4 2011-01-04      1   0     1       0       2          1
1
4        5 2011-01-05      1   0     1       0       3          1
1

        temp     atemp      hum  windspeed  casual  registered   cnt
0  14.110847  18.18125  80.5833  10.749882     331         654   985
1  14.902598  17.68695  69.6087  16.652113     131         670   801
2   8.050924   9.47025  43.7273  16.636703     120        1229  1349
3   8.200000  10.60610  59.0435  10.739832     108        1454  1562
4   9.305237  11.46350  43.6957  12.522300      82        1518  1600
```

```
day_df.describe()
```

```
          instant          yr        mnth        temp       atemp
hum  \
count  731.000000  731.000000  731.000000  731.000000  731.000000
731.000000
mean   366.000000    0.500684    6.519836   20.310776   23.717699
62.789406
std    211.165812    0.500342    3.451913    7.505091    8.148059
14.242910
min      1.000000    0.000000    1.000000    2.424346    3.953480
0.000000
25%    183.500000    0.000000    4.000000   13.820424   16.892125
52.000000
50%    366.000000    1.000000    7.000000   20.431653   24.336650
62.666700
75%    548.500000    1.000000   10.000000   26.872076   30.430100
```

```
73.020850
max    731.000000    1.000000   12.000000   35.328347   42.044800
97.250000
```

```
        windspeed       casual   registered          cnt
count  731.000000   731.000000   731.000000   731.000000
mean    12.762576   848.176471  3656.172367  4504.348837
std      5.192357   686.622488  1560.256377  1937.211452
min      1.500244     2.000000    20.000000    22.000000
25%      9.041650   315.500000  2497.000000  3152.000000
50%     12.125325   713.000000  3662.000000  4548.000000
75%     15.625371  1096.000000  4776.500000  5956.000000
max     34.000021  3410.000000  6946.000000  8714.000000
```

```
for i in col:
    print("Name of {} col".format(i)) #Name of Col
    print("No. of NUnique", day_df[i].nunique()) #Total Nunique Values
    print("Unique Values", day_df[i].unique())# All unique vales
    print('*'*30) # to make differnce i each col
    print()
    print()
```

```
Name of season col
No. of NUnique 4
Unique Values [1, 2, 3, 4]
Categories (4, int64): [1, 2, 3, 4]
******************************


Name of holiday col
No. of NUnique 2
Unique Values [0, 1]
Categories (2, int64): [0, 1]
******************************


Name of weekday col
No. of NUnique 7
Unique Values [6, 0, 1, 2, 3, 4, 5]
Categories (7, int64): [0, 1, 2, 3, 4, 5, 6]
******************************


Name of workingday col
No. of NUnique 2
Unique Values [0, 1]
Categories (2, int64): [0, 1]
******************************


Name of weathersit col
```

```
No. of NUnique 3
Unique Values [2, 1, 3]
Categories (3, int64): [1, 2, 3]
******************************
```

Exploratory Data Analysis (EDA)

Explore day_df

```
day_df.describe(include='all')
```

```
<ipython-input-40-a3f85d0fb009>:1: FutureWarning: Treating datetime
data as categorical rather than numeric in `.describe` is deprecated
and will be removed in a future version of pandas. Specify
`datetime_is_numeric=True` to silence this warning and adopt the
future behavior now.
  day_df.describe(include='all')
```

```
           instant                dteday  season          yr
mnth  \
count   731.000000                   731   731.0  731.000000
731.000000
unique         NaN                   731     4.0         NaN
NaN
top            NaN   2011-01-01 00:00:00     3.0         NaN
NaN
freq           NaN                     1   188.0         NaN
NaN
first          NaN   2011-01-01 00:00:00     NaN         NaN
NaN
last           NaN   2012-12-31 00:00:00     NaN         NaN
NaN
mean    366.000000                   NaN     NaN    0.500684
6.519836
std     211.165812                   NaN     NaN    0.500342
3.451913
min       1.000000                   NaN     NaN    0.000000
1.000000
25%     183.500000                   NaN     NaN    0.000000
4.000000
50%     366.000000                   NaN     NaN    1.000000
7.000000
75%     548.500000                   NaN     NaN    1.000000
10.000000
max     731.000000                   NaN     NaN    1.000000
12.000000

        holiday  weekday  workingday  weathersit        temp
```

```
atemp  \
count     731.0    731.0       731.0       731.0  731.000000
731.000000
unique      2.0      7.0         2.0         3.0         NaN
NaN
top         0.0      0.0         1.0         1.0         NaN
NaN
freq      710.0    105.0       500.0       463.0         NaN
NaN
first       NaN      NaN         NaN         NaN         NaN
NaN
last        NaN      NaN         NaN         NaN         NaN
NaN
mean        NaN      NaN         NaN         NaN   20.310776
23.717699
std         NaN      NaN         NaN         NaN    7.505091
8.148059
min         NaN      NaN         NaN         NaN    2.424346
3.953480
25%         NaN      NaN         NaN         NaN   13.820424
16.892125
50%         NaN      NaN         NaN         NaN   20.431653
24.336650
75%         NaN      NaN         NaN         NaN   26.872076
30.430100
max         NaN      NaN         NaN         NaN   35.328347
42.044800

                hum   windspeed       casual   registered          cnt

count   731.000000  731.000000   731.000000   731.000000   731.000000

unique         NaN         NaN          NaN          NaN          NaN

top            NaN         NaN          NaN          NaN          NaN

freq           NaN         NaN          NaN          NaN          NaN

first          NaN         NaN          NaN          NaN          NaN

last           NaN         NaN          NaN          NaN          NaN

mean     62.789406   12.762576   848.176471  3656.172367  4504.348837

std      14.242910    5.192357   686.622488  1560.256377  1937.211452

min       0.000000    1.500244     2.000000    20.000000    22.000000

25%      52.000000    9.041650   315.500000  2497.000000  3152.000000
```

```
50%      62.666700   12.125325   713.000000  3662.000000  4548.000000
75%      73.020850   15.625371  1096.000000  4776.500000  5956.000000
max      97.250000   34.000021  3410.000000  6946.000000  8714.000000
```

Mengurutkan Jumlah Peminjam Sepeda berdasarkan tahunnya

```
day_df.groupby(by="yr").cnt.nunique().sort_values(ascending=False)
```

```
yr
1    356
0    352
Name: cnt, dtype: int64
```

Mengurutkan Jumlah Peminjam Sepeda berdasarkan bulan

```
day_df.groupby(by="mnth").cnt.nunique().sort_values(ascending=False)
```

```
mnth
1     62
5     62
7     62
8     62
10    62
12    62
3     60
4     60
6     60
9     60
11    60
2     57
Name: cnt, dtype: int64
```

Mengurutkan Jumlah Peminjam Sepeda berdasarkan Musim

```
day_df.groupby(by="season").cnt.nunique().sort_values(ascending=False)
```

```
season
3    188
2    182
1    179
4    176
Name: cnt, dtype: int64
```

Mencari Kecepatan maksimal,minimal,dan rata-rata dari angin berdasarkan tahun

```
day_df.groupby(by="yr").agg({
    "windspeed": ["max", "min", "mean"]
})
```

yr	windspeed		
	max	min	mean
0	34.000021	1.500244	12.823977
1	29.584721	3.125550	12.701344

Mencari maksimal,minimal,dan rata-rata dari temperature berdasarkan tahun

```
day_df.groupby(by="yr").agg({
    "temp": ["max", "min", "mean"]
})
```

yr	temp		
	max	min	mean
0	34.815847	2.424346	19.953263
1	35.328347	4.407500	20.667313

```
day_df.groupby(by="season").agg({
    "temp": ["max", "min", "mean"]
})
```

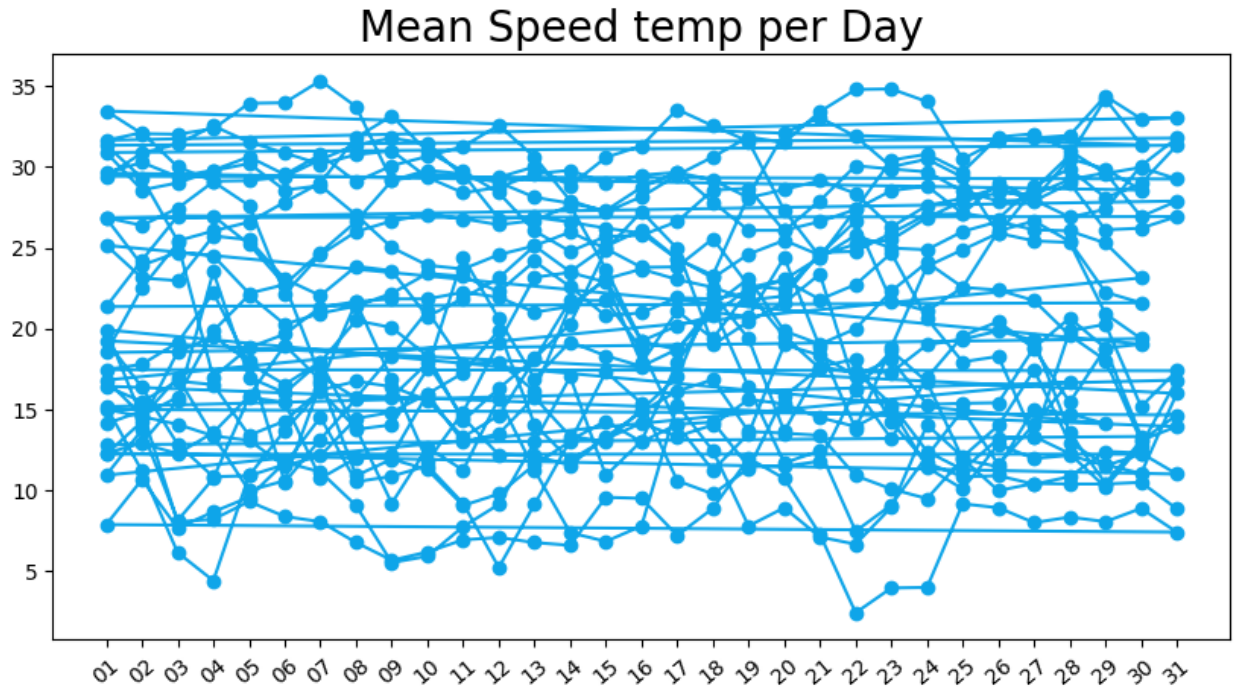
season	temp		
	max	min	mean
1	23.472500	2.424346	12.207650
2	33.141653	10.374763	22.320611
3	35.328347	19.235847	28.958682
4	26.957500	9.054153	17.339148

```
day_df.groupby(by="season").agg({
    "windspeed": ["max", "min", "mean"]
})
```

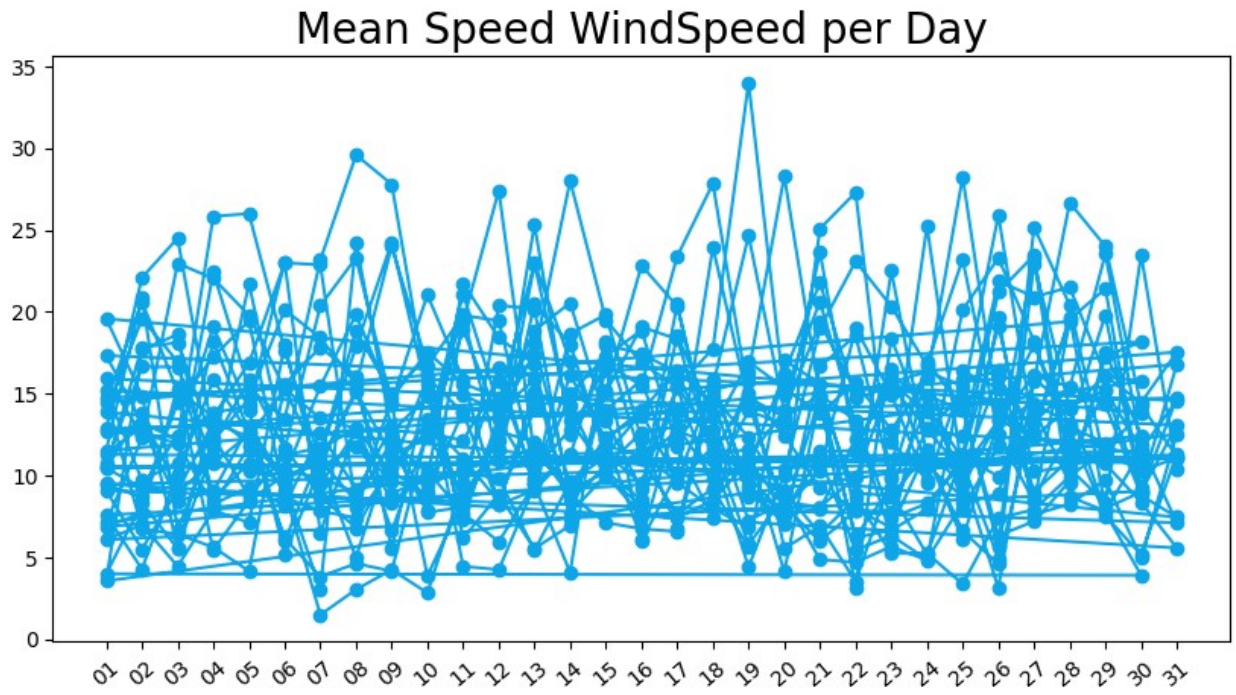
season	windspeed		
	max	min	mean
1	34.000021	3.042356	14.373984
2	26.000489	4.417256	13.634978
3	25.166339	4.292744	11.530366
4	28.292425	1.500244	11.523637

```
day_temp = day_df.resample(rule='D', on='dteday').agg({
    "temp": "mean"
})

day_temp.index = day_temp.index.strftime('%d')
```

```
plt.figure(figsize=(10, 5))
plt.plot(day_temp.index, day_temp['temp'], marker='o',
color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Speed temp per Day', loc='center', fontsize=20)
plt.show()
```




```
day_wind = day_df.resample(rule='D', on='dteday').agg({
    "windspeed": "mean"
})

day_wind.index = day_wind.index.strftime('%d')

plt.figure(figsize=(10, 5))
plt.plot(day_wind.index, day_wind['windspeed'], marker='o',
color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Speed WindSpeed per Day', loc='center',
fontsize=20)
plt.show()
```



Visualization & Explanatory Analysis

Pertanyaan 1:

Berapa Rata Rata Kecepatan Angin Pada tahun 2011,Berapa Suhu Temperature pada tahun 2011

```
day_df.groupby(by="yr").agg({
    "windspeed": ["max", "min", "mean"]
})
```

	windspeed		
	max	min	mean
yr			
0	34.000021	1.500244	12.823977
1	29.584721	3.125550	12.701344

```
day_df[day_df.mnth == 4]
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
90	91	2011-04-01	2	0	4	0	5	1	
91	92	2011-04-02	2	0	4	0	6	0	
92	93	2011-04-03	2	0	4	0	0	0	
93	94	2011-04-04	2	0	4	0	1	1	
94	95	2011-04-05	2	0	4	0	2	1	
95	96	2011-04-06	2	0	4	0	3	1	
96	97	2011-04-07	2	0	4	0	4	1	
97	98	2011-04-08	2	0	4	0	5	1	
98	99	2011-04-09	2	0	4	0	6	0	

99	100	2011-04-10	2	0	4	0	0	0
100	101	2011-04-11	2	0	4	0	1	1
101	102	2011-04-12	2	0	4	0	2	1
102	103	2011-04-13	2	0	4	0	3	1
103	104	2011-04-14	2	0	4	0	4	1
104	105	2011-04-15	2	0	4	1	5	0
105	106	2011-04-16	2	0	4	0	6	0
106	107	2011-04-17	2	0	4	0	0	0
107	108	2011-04-18	2	0	4	0	1	1
108	109	2011-04-19	2	0	4	0	2	1
109	110	2011-04-20	2	0	4	0	3	1
110	111	2011-04-21	2	0	4	0	4	1
111	112	2011-04-22	2	0	4	0	5	1
112	113	2011-04-23	2	0	4	0	6	0
113	114	2011-04-24	2	0	4	0	0	0
114	115	2011-04-25	2	0	4	0	1	1
115	116	2011-04-26	2	0	4	0	2	1
116	117	2011-04-27	2	0	4	0	3	1
117	118	2011-04-28	2	0	4	0	4	1
118	119	2011-04-29	2	0	4	0	5	1
119	120	2011-04-30	2	0	4	0	6	0
456	457	2012-04-01	2	1	4	0	0	0
457	458	2012-04-02	2	1	4	0	1	1
458	459	2012-04-03	2	1	4	0	2	1
459	460	2012-04-04	2	1	4	0	3	1
460	461	2012-04-05	2	1	4	0	4	1
461	462	2012-04-06	2	1	4	0	5	1
462	463	2012-04-07	2	1	4	0	6	0
463	464	2012-04-08	2	1	4	0	0	0
464	465	2012-04-09	2	1	4	0	1	1
465	466	2012-04-10	2	1	4	0	2	1
466	467	2012-04-11	2	1	4	0	3	1
467	468	2012-04-12	2	1	4	0	4	1
468	469	2012-04-13	2	1	4	0	5	1
469	470	2012-04-14	2	1	4	0	6	0
470	471	2012-04-15	2	1	4	0	0	0
471	472	2012-04-16	2	1	4	1	1	0
472	473	2012-04-17	2	1	4	0	2	1
473	474	2012-04-18	2	1	4	0	3	1
474	475	2012-04-19	2	1	4	0	4	1
475	476	2012-04-20	2	1	4	0	5	1
476	477	2012-04-21	2	1	4	0	6	0
477	478	2012-04-22	2	1	4	0	0	0
478	479	2012-04-23	2	1	4	0	1	1
479	480	2012-04-24	2	1	4	0	2	1
480	481	2012-04-25	2	1	4	0	3	1
481	482	2012-04-26	2	1	4	0	4	1
482	483	2012-04-27	2	1	4	0	5	1
483	484	2012-04-28	2	1	4	0	6	0

```
484      485 2012-04-29      2   1     4       0       0          0
485      486 2012-04-30      2   1     4       0       1          1

    weathersit       temp     atemp      hum  windspeed  casual
registered  \
90           2  12.300000  14.17270  68.6250  17.333436     307
1920
91           2  12.915000  15.78185  65.3750  13.208782     898
1354
92           1  15.511653  18.93835  48.0000  12.208271    1651
1598
93           1  23.506653  27.14645  42.6250  25.833257     734
2381
94           2  16.980847  19.91750  64.2083  26.000489     167
1628
95           1  16.024153  19.38040  47.0833  17.625221     413
2395
96           1  17.937500  21.68480  60.2917  10.874904     571
2570
97           2  13.769153  16.22395  83.6250  15.208464     172
1299
98           2  14.042500  17.07645  87.7500   8.916561     879
1576
99           2  17.493347  21.33685  85.7500   9.833389    1188
1707
100          2  24.421732  28.26085  71.6956  21.739758     855
2493
101          2  20.602500  24.65270  73.9167  18.416893     257
1777
102          2  16.912500  20.86415  81.9167  16.791339     209
1953
103          1  19.167500  23.13710  54.0417   7.416900     529
2738
104          1  18.313347  22.09565  67.1250  15.167125     642
2484
105          3  17.664153  21.27460  88.8333  22.834136     121
674
106          1  18.723347  22.28480  47.9583  20.334232    1558
2186
107          1  21.012500  25.15730  54.2500  10.958989     669
2760
108          2  20.739153  24.46290  66.5833  10.584057     409
2795
109          1  24.395000  28.21960  61.4167  16.208975     613
3331
110          1  18.825847  22.69460  40.7083  21.792286     745
3444
111          2  13.803347  16.09770  72.9583  14.707907     177
1506
```

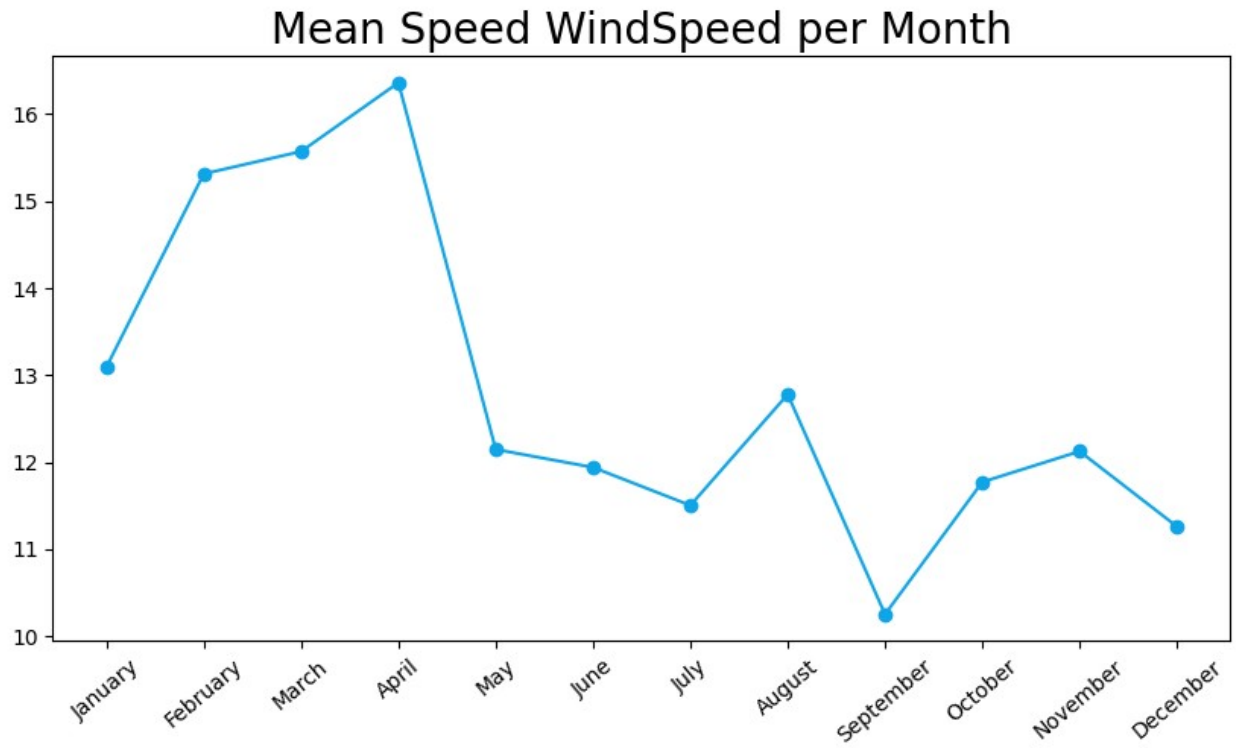
```
112          2  18.860000  22.50605  88.7917  15.458575    1462
2574
113          2  23.848347  27.58815  81.0833  12.875725    1710
2481
114          1  24.873347  28.72500  77.6667  12.417311     773
3300
115          1  25.898347  29.70415  72.9167  21.875500     678
3722
116          2  25.420000  28.75710  83.5417  20.917400     547
3325
117          2  25.317500  28.94645  70.0833  21.500836     569
3489
118          1  20.910000  24.87315  45.7083  16.084221     878
3717
119          1  19.372500  23.20105  50.3333  15.750025    1965
3347
456          2  17.459153  20.86435  67.6250  11.541889    2347
3694
457          1  17.790433  21.37565  50.4348  20.913313    1208
4728
458          1  19.133347  23.07415  39.6667   6.708911    1348
5424
459          1  22.208347  26.67250  46.9583  12.125325    1058
5378
460          1  17.835000  21.55815  37.4167  14.708443    1192
5265
461          1  16.536653  19.53835  37.7083  20.125996    1807
4653
462          1  17.937500  21.30645  25.4167  18.416357    3252
3605
463          1  20.500000  24.62125  27.5833  15.583932    2230
2939
464          1  20.055847  23.83190  31.7500  23.999132     905
4680
465          1  18.313347  21.81165  43.5000  16.708125     819
5099
466          1  14.296536  16.86370  46.9565  19.783358     482
4380
467          1  16.297500  19.38020  46.6250  19.458743     663
4746
468          1  18.142500  21.59040  40.8333  10.416557    1252
5146
469          1  20.295000  24.39980  50.2917  12.791439    2795
4665
470          1  24.873347  28.69375  50.7917  15.083643    2846
4286
471          1  27.230847  30.74625  56.1667  19.083543    1198
5172
472          1  24.941653  29.92435  39.0417  18.333143     989
```

```
5702
473          2  18.996653  22.85190  56.9167  11.250104     347
4020
474          1  20.431653  24.65230  61.2500   4.417256     846
5719
475          1  21.593347  25.78875  69.4583  10.041357    1340
5950
476          1  23.370000  27.14605  68.2917  19.000329    2541
4083
477          3  16.263347  19.47520  83.5417  23.084582     120
907
478          2  13.188347  15.05625  76.6667  20.334232     195
3019
479          1  16.946653  20.26415  45.4167  16.708661     518
5115
480          1  19.543347  23.51585  42.7917   7.959064     655
5541
481          2  20.431653  24.17915  75.6667  11.833875     475
4551
482          1  18.757500  22.63185  40.0833  23.291411    1014
5219
483          2  15.443347  18.87520  48.9583   8.708325    1120
3100
484          1  18.791653  22.50605  58.7083   7.832836    2229
4075
485          2  19.030847  22.88480  57.0000  11.499746     665
4907

      cnt
90   2227
91   2252
92   3249
93   3115
94   1795
95   2808
96   3141
97   1471
98   2455
99   2895
100  3348
101  2034
102  2162
103  3267
104  3126
105   795
106  3744
107  3429
108  3204
109  3944
```

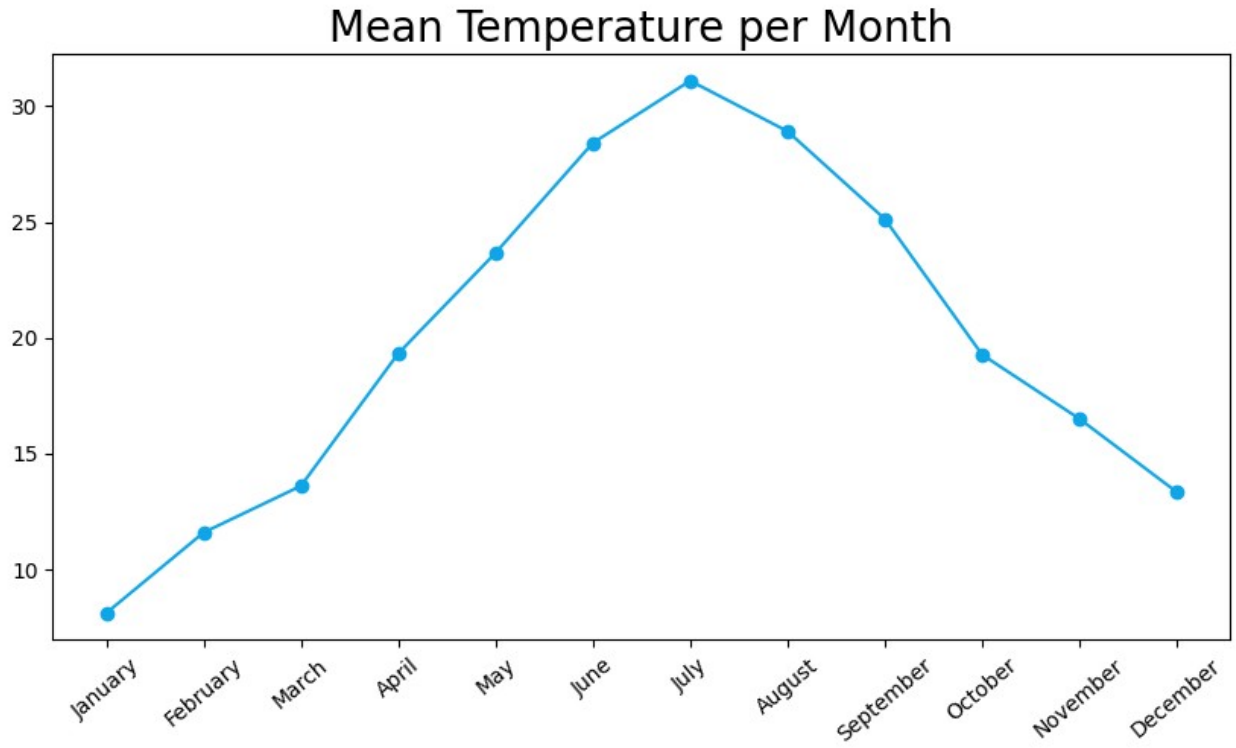
```
110  4189
111  1683
112  4036
113  4191
114  4073
115  4400
116  3872
117  4058
118  4595
119  5312
456  6041
457  5936
458  6772
459  6436
460  6457
461  6460
462  6857
463  5169
464  5585
465  5918
466  4862
467  5409
468  6398
469  7460
470  7132
471  6370
472  6691
473  4367
474  6565
475  7290
476  6624
477  1027
478  3214
479  5633
480  6196
481  5026
482  6233
483  4220
484  6304
485  5572
```

```
year_windspeed = day_df[day_df.dteday.dt.year == 2011]
year_windspeed = year_windspeed.resample(rule='M', on='dteday').agg({
    "windspeed": "mean"
})

year_windspeed.index = year_windspeed.index.strftime('%B')

plt.figure(figsize=(10, 5))
plt.plot(year_windspeed.index, year_windspeed['windspeed'],
```

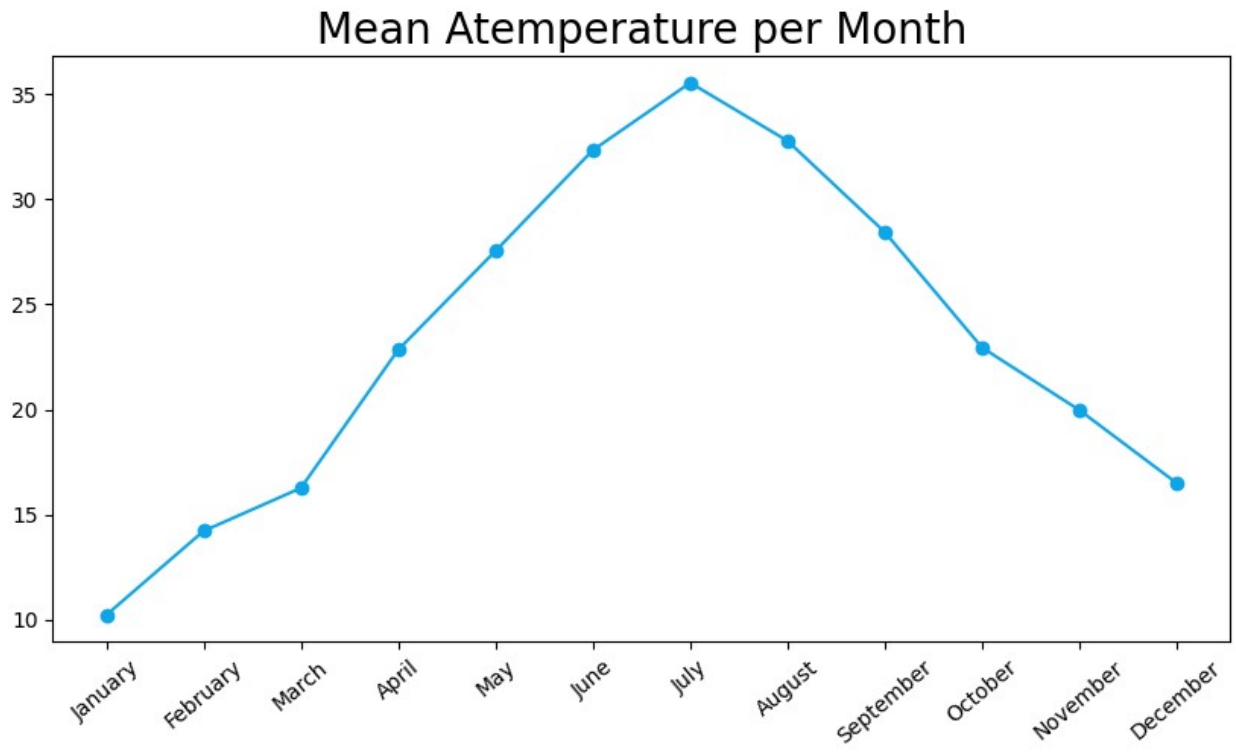
```
marker='o', color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Speed WindSpeed per Month', loc='center',
fontsize=20)
plt.show()
```



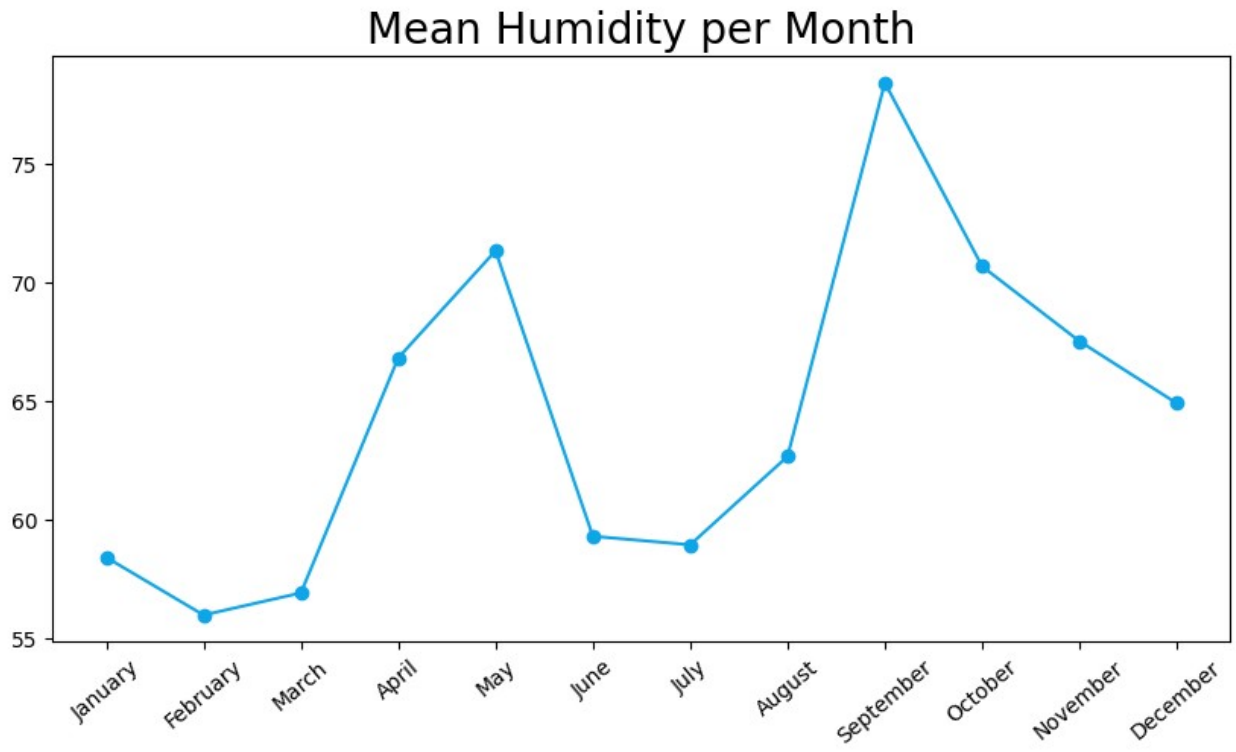

```
year_temp = day_df[day_df.dteday.dt.year == 2011]
year_temp = year_temp.resample(rule='M', on='dteday').agg({
    "temp": "mean"
})

year_temp.index = year_temp.index.strftime('%B')

plt.figure(figsize=(10, 5))
plt.plot(year_temp.index, year_temp['temp'], marker='o',
color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Temperature per Month', loc='center',
fontsize=20)
plt.show()
```




```
year_atemp = day_df[day_df.dteday.dt.year == 2011]
year_atemp = year_atemp.resample(rule='M', on='dteday').agg({
    "atemp": "mean"
})

year_atemp.index = year_atemp.index.strftime('%B')

plt.figure(figsize=(10, 5))
plt.plot(year_atemp.index, year_atemp['atemp'], marker='o',
color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Atemperature per Month', loc='center',
fontsize=20)
plt.show()
```




```
year_hum = day_df[day_df.dteday.dt.year == 2011]
year_hum = year_hum.resample(rule='M', on='dteday').agg({
    "hum": "mean"
})

year_hum.index = year_hum.index.strftime('%B')

plt.figure(figsize=(10, 5))
plt.plot(year_hum.index, year_hum['hum'], marker='o', color='#0ea5e9')
plt.xticks(rotation=40)
plt.title(label='Mean Humidity per Month', loc='center', fontsize=20)
plt.show()
```




```
day_df[day_df.mnth == 4]
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
90	91	2011-04-01	2	0	4	0	5	1	
91	92	2011-04-02	2	0	4	0	6	0	
92	93	2011-04-03	2	0	4	0	0	0	
93	94	2011-04-04	2	0	4	0	1	1	
94	95	2011-04-05	2	0	4	0	2	1	
95	96	2011-04-06	2	0	4	0	3	1	
96	97	2011-04-07	2	0	4	0	4	1	
97	98	2011-04-08	2	0	4	0	5	1	
98	99	2011-04-09	2	0	4	0	6	0	
99	100	2011-04-10	2	0	4	0	0	0	
100	101	2011-04-11	2	0	4	0	1	1	
101	102	2011-04-12	2	0	4	0	2	1	
102	103	2011-04-13	2	0	4	0	3	1	
103	104	2011-04-14	2	0	4	0	4	1	
104	105	2011-04-15	2	0	4	1	5	0	
105	106	2011-04-16	2	0	4	0	6	0	
106	107	2011-04-17	2	0	4	0	0	0	
107	108	2011-04-18	2	0	4	0	1	1	
108	109	2011-04-19	2	0	4	0	2	1	
109	110	2011-04-20	2	0	4	0	3	1	
110	111	2011-04-21	2	0	4	0	4	1	
111	112	2011-04-22	2	0	4	0	5	1	
112	113	2011-04-23	2	0	4	0	6	0	

113	114	2011-04-24	2	0	4	0	0	0
114	115	2011-04-25	2	0	4	0	1	1
115	116	2011-04-26	2	0	4	0	2	1
116	117	2011-04-27	2	0	4	0	3	1
117	118	2011-04-28	2	0	4	0	4	1
118	119	2011-04-29	2	0	4	0	5	1
119	120	2011-04-30	2	0	4	0	6	0
456	457	2012-04-01	2	1	4	0	0	0
457	458	2012-04-02	2	1	4	0	1	1
458	459	2012-04-03	2	1	4	0	2	1
459	460	2012-04-04	2	1	4	0	3	1
460	461	2012-04-05	2	1	4	0	4	1
461	462	2012-04-06	2	1	4	0	5	1
462	463	2012-04-07	2	1	4	0	6	0
463	464	2012-04-08	2	1	4	0	0	0
464	465	2012-04-09	2	1	4	0	1	1
465	466	2012-04-10	2	1	4	0	2	1
466	467	2012-04-11	2	1	4	0	3	1
467	468	2012-04-12	2	1	4	0	4	1
468	469	2012-04-13	2	1	4	0	5	1
469	470	2012-04-14	2	1	4	0	6	0
470	471	2012-04-15	2	1	4	0	0	0
471	472	2012-04-16	2	1	4	1	1	0
472	473	2012-04-17	2	1	4	0	2	1
473	474	2012-04-18	2	1	4	0	3	1
474	475	2012-04-19	2	1	4	0	4	1
475	476	2012-04-20	2	1	4	0	5	1
476	477	2012-04-21	2	1	4	0	6	0
477	478	2012-04-22	2	1	4	0	0	0
478	479	2012-04-23	2	1	4	0	1	1
479	480	2012-04-24	2	1	4	0	2	1
480	481	2012-04-25	2	1	4	0	3	1
481	482	2012-04-26	2	1	4	0	4	1
482	483	2012-04-27	2	1	4	0	5	1
483	484	2012-04-28	2	1	4	0	6	0
484	485	2012-04-29	2	1	4	0	0	0
485	486	2012-04-30	2	1	4	0	1	1

	weathersit	temp	atemp	hum	windspeed	casual	registered	\
90	2	12.300000	14.17270	68.6250	17.333436	307	1920	
91	2	12.915000	15.78185	65.3750	13.208782	898	1354	
92	1	15.511653	18.93835	48.0000	12.208271	1651	1598	
93	1	23.506653	27.14645	42.6250	25.833257	734	2381	
94	2	16.980847	19.91750	64.2083	26.000489	167		

```
1628
95           1  16.024153  19.38040  47.0833  17.625221     413
2395
96           1  17.937500  21.68480  60.2917  10.874904     571
2570
97           2  13.769153  16.22395  83.6250  15.208464     172
1299
98           2  14.042500  17.07645  87.7500   8.916561     879
1576
99           2  17.493347  21.33685  85.7500   9.833389    1188
1707
100          2  24.421732  28.26085  71.6956  21.739758     855
2493
101          2  20.602500  24.65270  73.9167  18.416893     257
1777
102          2  16.912500  20.86415  81.9167  16.791339     209
1953
103          1  19.167500  23.13710  54.0417   7.416900     529
2738
104          1  18.313347  22.09565  67.1250  15.167125     642
2484
105          3  17.664153  21.27460  88.8333  22.834136     121
674
106          1  18.723347  22.28480  47.9583  20.334232    1558
2186
107          1  21.012500  25.15730  54.2500  10.958989     669
2760
108          2  20.739153  24.46290  66.5833  10.584057     409
2795
109          1  24.395000  28.21960  61.4167  16.208975     613
3331
110          1  18.825847  22.69460  40.7083  21.792286     745
3444
111          2  13.803347  16.09770  72.9583  14.707907     177
1506
112          2  18.860000  22.50605  88.7917  15.458575    1462
2574
113          2  23.848347  27.58815  81.0833  12.875725    1710
2481
114          1  24.873347  28.72500  77.6667  12.417311     773
3300
115          1  25.898347  29.70415  72.9167  21.875500     678
3722
116          2  25.420000  28.75710  83.5417  20.917400     547
3325
117          2  25.317500  28.94645  70.0833  21.500836     569
3489
118          1  20.910000  24.87315  45.7083  16.084221     878
3717
```

```
119          1  19.372500  23.20105  50.3333  15.750025    1965
3347
456          2  17.459153  20.86435  67.6250  11.541889    2347
3694
457          1  17.790433  21.37565  50.4348  20.913313    1208
4728
458          1  19.133347  23.07415  39.6667   6.708911    1348
5424
459          1  22.208347  26.67250  46.9583  12.125325    1058
5378
460          1  17.835000  21.55815  37.4167  14.708443    1192
5265
461          1  16.536653  19.53835  37.7083  20.125996    1807
4653
462          1  17.937500  21.30645  25.4167  18.416357    3252
3605
463          1  20.500000  24.62125  27.5833  15.583932    2230
2939
464          1  20.055847  23.83190  31.7500  23.999132     905
4680
465          1  18.313347  21.81165  43.5000  16.708125     819
5099
466          1  14.296536  16.86370  46.9565  19.783358     482
4380
467          1  16.297500  19.38020  46.6250  19.458743     663
4746
468          1  18.142500  21.59040  40.8333  10.416557    1252
5146
469          1  20.295000  24.39980  50.2917  12.791439    2795
4665
470          1  24.873347  28.69375  50.7917  15.083643    2846
4286
471          1  27.230847  30.74625  56.1667  19.083543    1198
5172
472          1  24.941653  29.92435  39.0417  18.333143     989
5702
473          2  18.996653  22.85190  56.9167  11.250104     347
4020
474          1  20.431653  24.65230  61.2500   4.417256     846
5719
475          1  21.593347  25.78875  69.4583  10.041357    1340
5950
476          1  23.370000  27.14605  68.2917  19.000329    2541
4083
477          3  16.263347  19.47520  83.5417  23.084582     120
907
478          2  13.188347  15.05625  76.6667  20.334232     195
3019
479          1  16.946653  20.26415  45.4167  16.708661     518
```

```
5115
480          1  19.543347  23.51585  42.7917   7.959064     655
5541
481          2  20.431653  24.17915  75.6667  11.833875     475
4551
482          1  18.757500  22.63185  40.0833  23.291411    1014
5219
483          2  15.443347  18.87520  48.9583   8.708325    1120
3100
484          1  18.791653  22.50605  58.7083   7.832836    2229
4075
485          2  19.030847  22.88480  57.0000  11.499746     665
4907
```

	cnt
90	2227
91	2252
92	3249
93	3115
94	1795
95	2808
96	3141
97	1471
98	2455
99	2895
100	3348
101	2034
102	2162
103	3267
104	3126
105	795
106	3744
107	3429
108	3204
109	3944
110	4189
111	1683
112	4036
113	4191
114	4073
115	4400
116	3872
117	4058
118	4595
119	5312
456	6041
457	5936
458	6772
459	6436

```
460  6457
461  6460
462  6857
463  5169
464  5585
465  5918
466  4862
467  5409
468  6398
469  7460
470  7132
471  6370
472  6691
473  4367
474  6565
475  7290
476  6624
477  1027
478  3214
479  5633
480  6196
481  5026
482  6233
483  4220
484  6304
485  5572
```

```
day_df[day_df.mnth == 7]
```

```
     instant     dteday season  yr  mnth holiday weekday workingday  \
181      182 2011-07-01      3   0     7       0       5          1
182      183 2011-07-02      3   0     7       0       6          0
183      184 2011-07-03      3   0     7       0       0          0
184      185 2011-07-04      3   0     7       1       1          0
185      186 2011-07-05      3   0     7       0       2          1
..       ...        ...    ...  ..   ...     ...     ...        ...
573      574 2012-07-27      3   1     7       0       5          1
574      575 2012-07-28      3   1     7       0       6          0
575      576 2012-07-29      3   1     7       0       0          0
576      577 2012-07-30      3   1     7       0       1          1
577      578 2012-07-31      3   1     7       0       2          1

    weathersit       temp     atemp      hum  windspeed  casual
registered  \
181          1  29.622500  32.60810  39.6250   6.874736    1246
4116
182          1  30.271653  33.36540  44.4583   7.709154    2204
2915
183          2  29.383347  33.42875  68.2500  15.333486    2282
2367
```

```
184          2  29.793347  33.27085  63.7917   5.459106    3065
2978
185          1  30.613347  34.81690  59.0417   8.459286    1031
3634
..         ...        ...       ...      ...        ...     ...
...
573          1  32.048347  36.71085  59.4583  10.250464    1259
5645
574          1  30.989153  34.88020  61.3333  10.542450    2234
4451
575          1  29.588347  33.39665  62.3750  11.416532    2153
4444
576          1  29.964153  34.24935  66.8750  10.292339    1040
6065
577          1  29.246653  33.14480  70.4167  11.083475     968
6248

      cnt
181  5362
182  5119
183  4649
184  6043
185  4665
..    ...
573  6904
574  6685
575  6597
576  7105
577  7216

[62 rows x 16 columns]
```

```
day_df.groupby(by="season").cnt.nunique().sort_values(ascending=False)
```

```
season
3    188
2    182
1    179
4    176
Name: cnt, dtype: int64
```

Pertanyaan 1:

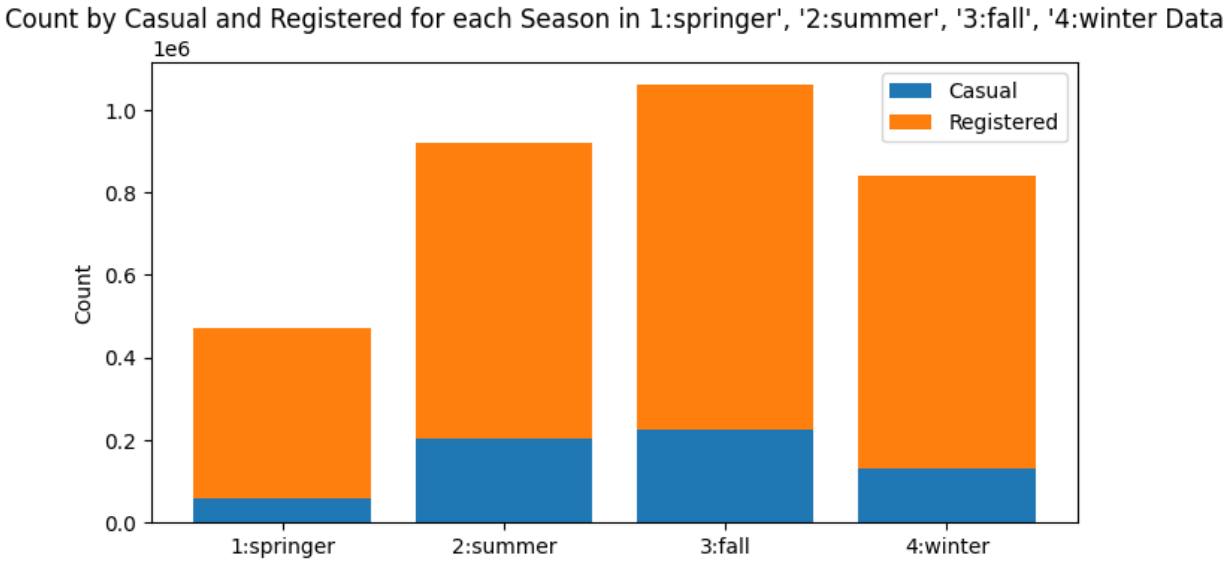
Berapa Rata Rata Kecepatan Angin Pada tahun 2011,Berapa Suhu Temperature pada tahun 2011

> Dari Data diatas menunjukan bahwa pada musim gugur banyak orang bersepeda hal tersebut menunjukan bahwa dengan suhu yang cukup panas dapat membakar banyak kalori sehingga banyak yang bersepeda pada bulan july dan bertepatan pada musim gugur

Dan pada data diatas yaitu ketika musim panas dan bertepatan pada bulan april banyak orang juga melakukan bersepeda karena angin dan suhu yang cukup baik untuk bersepeda pada musim tersebut

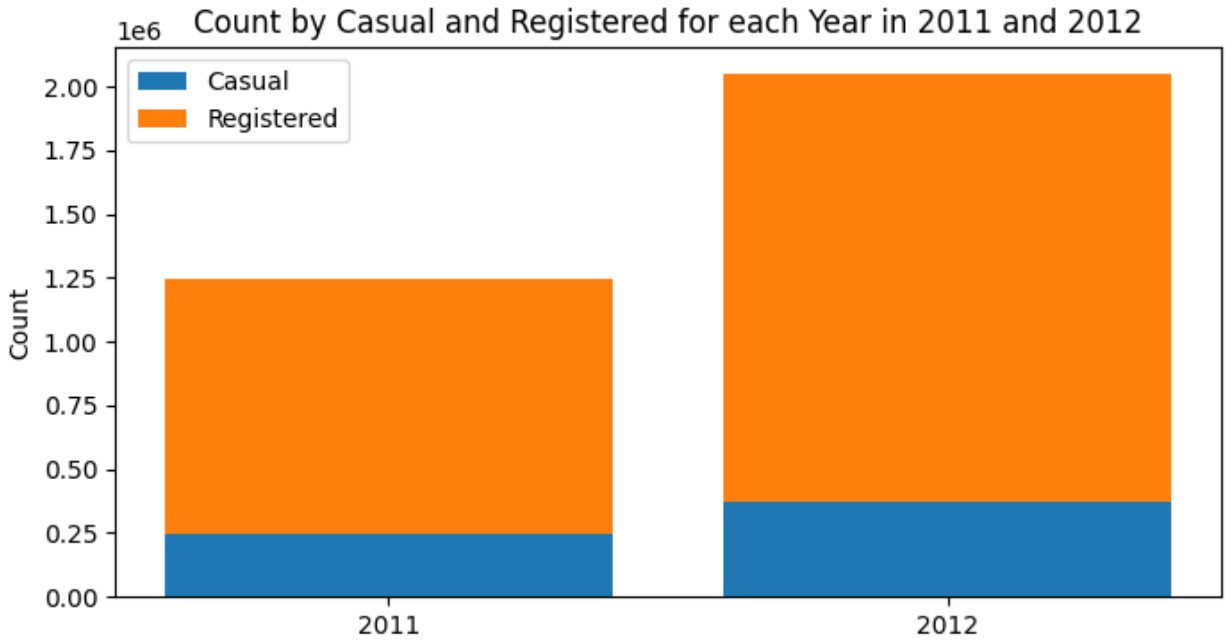
Pertanyaan 2: Berapa jumlah penyewa sepeda yang telah terdaftar dan yang belum terdaftar di tiap tahun

```
plt.figure(figsize=(8,4))
p1 = plt.bar(day_df['season'].unique(),
             day_df.groupby(['season'])['casual'].sum())
p2 = plt.bar(day_df['season'].unique(),  # the x locations for the
groups
             day_df.groupby(['season'])['registered'].sum(), # Count
of Registered per season
             bottom = day_df.groupby(['season'])['casual'].sum()) #
Count of casual per season

plt.ylabel('Count')
plt.title("Count by Casual and Registered for each Season in
1:springer', '2:summer', '3:fall', '4:winter Data")
plt.xticks(day_df['season'].unique(), ('1:springer', '2:summer',
'3:fall', '4:winter')) # Name of unique values in columns
plt.legend((p1[0], p2[0]), ('Casual', 'Registered')) #setting legends
as per target
plt.show()
```




```
plt.figure(figsize=(8,4))
p1 = plt.bar(day_df['yr'].unique(),
             day_df.groupby(['yr'])['casual'].sum())
p2 = plt.bar(day_df['yr'].unique(),  # the x locations for the groups
             day_df.groupby(['yr'])['registered'].sum(), # Count of
```

```
Registered per season
             bottom = day_df.groupby(['yr'])['casual'].sum()) # Count
of casual per season

plt.ylabel('Count')
plt.title("Count by Casual and Registered for each Year in 2011 and
2012")
plt.xticks(day_df['yr'].unique(), ('2011','2012')) # Name of unique
values in columns
plt.legend((p1[0], p2[0]), ('Casual', 'Registered')) #setting legends
as per target
plt.show()
```




```
day_df.groupby(by="yr").casual.nunique().sort_values(ascending=False)

yr
1    332
0    328
Name: casual, dtype: int64
```

Pertanyaan 2

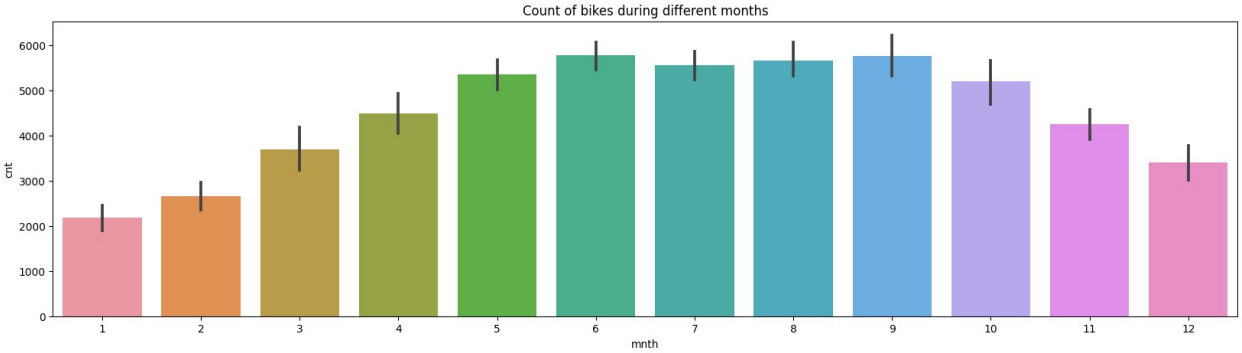
Berapa jumlah penyewa sepeda yang telah terdaftar dan yang belum terdaftar di tiap tahun

> Terlihat pada data diatas menunjukan bahwa jumlah penyewa sepeda terus meningkat setiap tahun pada tahun 2011 jumlah penyewa sepeda terlihat cukup banyak yaitu sekitar 328 jumlah yang menyewa sepeda kemudian pada tahun berikutnya terjadi

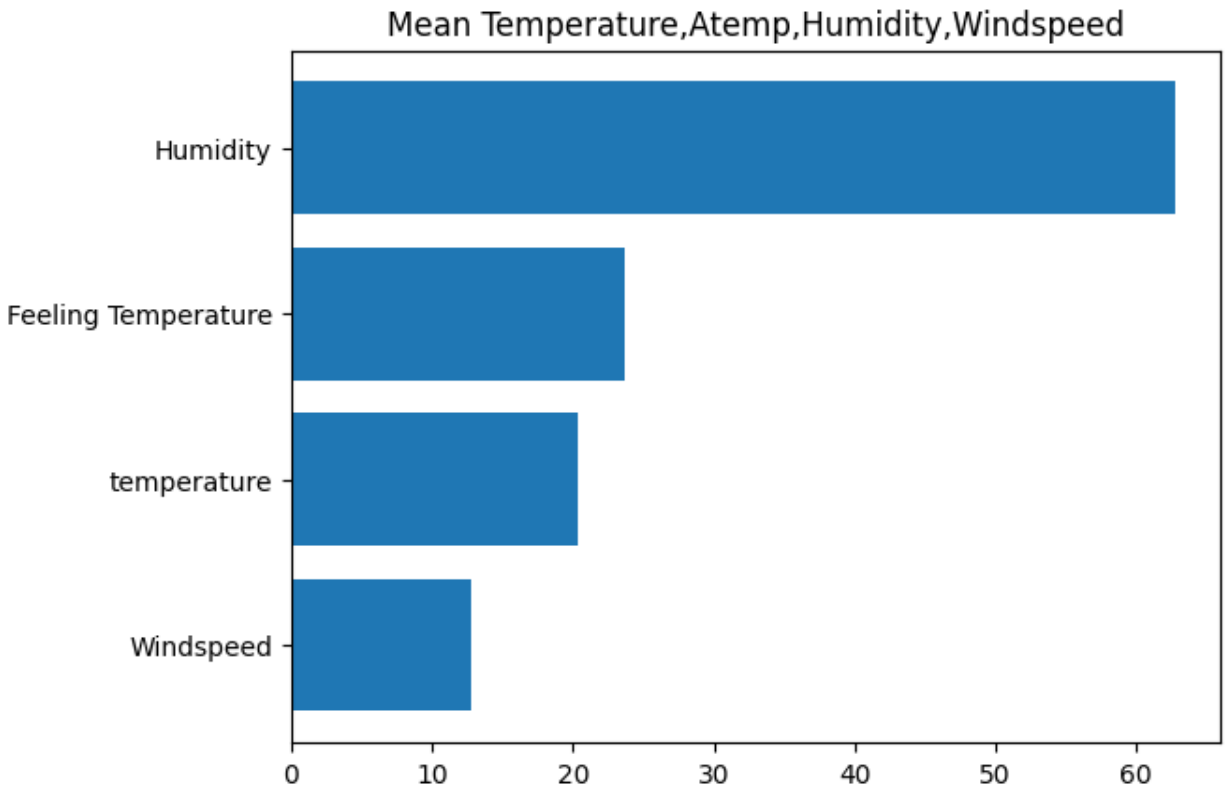
peningkatan yaitu sekitar 328 orang yang menyewa sepeda hal ini menunjukan bahwa jumlah peminat bersepada terus meningkat setiap tahunnya.

```
fig, ax = plt.subplots(figsize=(20,5))
sns.barplot(data=day_df, x='mnth', y='cnt', ax=ax)
ax.set(title='Count of bikes during different months')
```

```
[Text(0.5, 1.0, 'Count of bikes during different months')]
```




```
rfm_df = day_df.groupby(by=['temp', 'atemp', 'hum',
'windspeed']).agg({
    "temp": "mean",
    "atemp": "mean",
    "hum": "mean",
    "windspeed" : "mean"
})
rfm_df.columns = ["temperature", "Feeling Temperature",
"Humidity","Windspeed"]

rfm_df = rfm_df.mean().sort_values(ascending=True)

plt.barh(rfm_df.index, rfm_df.values)
plt.title(label="Mean Temperature,Atemp,Humidity,Windspeed")
plt.xlabel("")
plt.ylabel("")
plt.show()
```



Conclusion

- Conclution pertanyaan 1 : Berapa Kecepatan Rata rata angin pada tahun 2011

  Pada data yang telah diatas bisa kita liat bahwa rata rata kecepatan angin tiap bulan berbeda-beda Tingkat paling tinggi yaitu pada bulan April yaitu ketika musim panas.Kemudian kecepatan angin kembali menurun dibulan berikutnya seiring bergantinya musim dan pada musim panas ini banyak orang yang menyewa sepeda pada bulan dan musim tersebut.

- conclution pertanyaan 2 : Berapa jumlah penyewa sepeda yang telah terdaftar dan yang belum terdaftar di tiap tahun

  Setiap Tahunnya penyewa sepeda memiliki peningkatan.Pada tahun 2011 banyak penyewa sepeda yang sudah menjadi member dan ada beberapa penyewa sepeda yang belum menjadi member.Kemudian ditahun selanjutnya pada tahun 2012 jumlah penyewa sepeda meningkat dan jumlah penyewa sudah menjadi member pun memiliki peningkatan