Google Sheets Visualization

Project Description

Find or create a simple data source and put the data in tabular format into a Google Sheet. Your data source must include at least 10 rows of data and at least 4 columns describing that data. For example, you could describe 10 activities you do in a day including what each activity is called, when you start each activity, when you finish each, and how happy you were when doing that activity. The grain of this data would be one row per activity.

Using the Charts feature of Google Sheets, create a new sheet in this same file, create a simple visualization of this data using a minimum of 2 colors, and be sure to label your data points.

When you are finished with your viz, please respond to these two prompts on another new sheet in this same file:

- List at least 3 skills you used to create your visualization
- Who is this visualization for, i.e. who is your audience?

Import Libraries

In [1]:

```
import numpy as np
from numpy import count_nonzero, median, mean
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import random

import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols

%matplotlib inline
#sets the default autosave frequency in seconds
%autosave 60
sns.set_style('dark')
sns.set(font_scale=1.2)

plt.rc('axes', titlesize=9)
plt.rc('axes', labelsize=14)
plt.rc('xtick', labelsize=12)
plt.rc('ytick', labelsize=12)

import warnings
warnings.filterwarnings('ignore')

pd.set_option('display.max_columns',None)
#pd.set_option('display.max_rows',None)
pd.set_option('display.width', 1000)
pd.set_option('display.float_format','{:.2f}'.format)

random.seed(0)
np.random.seed(0)
np.set_printoptions(suppress=True)
```

```
Autosaving every 60 seconds
```

Exploratory Data Analysis

In [2]:
```
df = pd.read_csv("pedalme_features.csv")
```

In [3]:
```
df2 = df.head(30)
df2
```

Out[3]:

	year	week	location	time	demand
0	2020	22	0	0	12
1	2020	22	1	0	2
2	2020	22	2	0	0
3	2020	22	3	0	1
4	2020	22	4	0	0
5	2020	22	5	0	0
6	2020	22	6	0	1
7	2020	22	7	0	1
8	2020	22	8	0	1
9	2020	22	9	0	0
10	2020	22	10	0	0
11	2020	22	11	0	0
12	2020	22	12	0	0
13	2020	22	13	0	4
14	2020	22	14	0	0
15	2020	23	0	1	119
16	2020	23	1	1	30
17	2020	23	2	1	3
18	2020	23	3	1	31
19	2020	23	4	1	24
20	2020	23	5	1	14
21	2020	23	6	1	15
22	2020	23	7	1	7
23	2020	23	8	1	9
24	2020	23	9	1	5
25	2020	23	10	1	38
26	2020	23	11	1	4
27	2020	23	12	1	3

	year	week	location	time	demand
28	2020	23	13	1	10
29	2020	23	14	1	3

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 540 entries, 0 to 539
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   year      540 non-null    int64
 1   week      540 non-null    int64
 2   location  540 non-null    int64
 3   time      540 non-null    int64
 4   demand    540 non-null    int64
dtypes: int64(5)
memory usage: 21.2 KB
```

In [5]:
```
df2.groupby("week").mean()
```

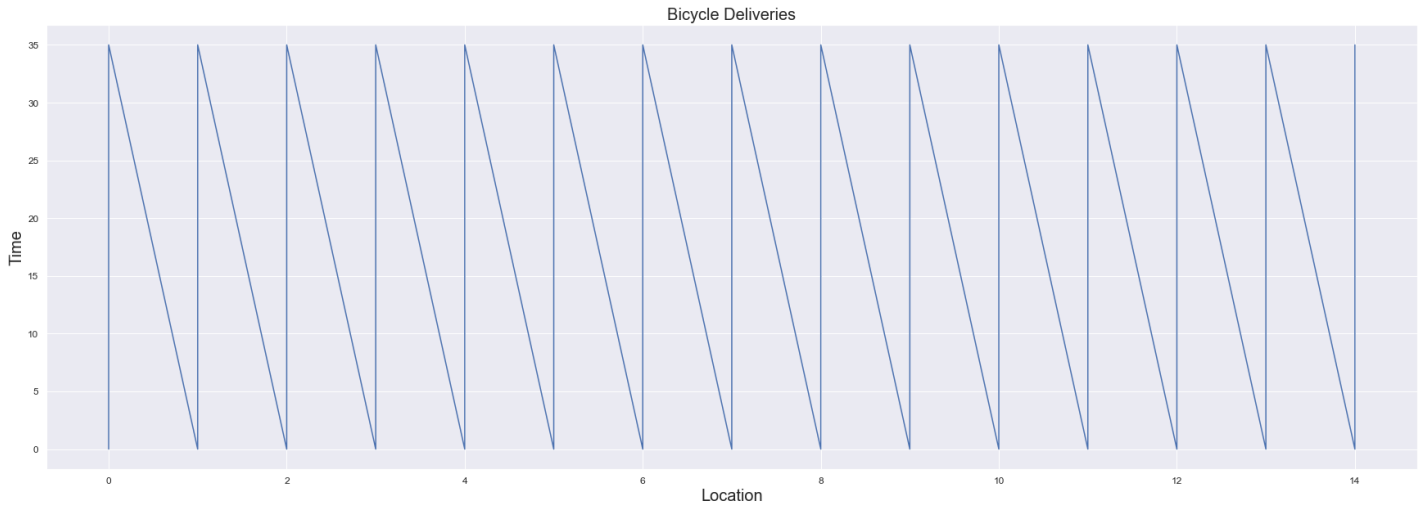
Out[5]:

week	year	location	time	demand
22	2020.00	7.00	0.00	1.47
23	2020.00	7.00	1.00	21.00

Data Visualization

Time-Series Analysis

In [6]:
```
fig = plt.figure(figsize=(30,10))
sns.lineplot(x=df.location,y=df.time, data=df, estimator=None)
plt.title("Bicycle Deliveries", fontsize=20)
plt.xlabel("Location", fontsize=20)
plt.ylabel("Time", fontsize=20)
#plt.legend(['week 22','week 23'])
plt.show()
```



List at least 3 skills you used to create your visualization

1. Programming
2. Data Analysis
3. Business Objectives

Who is this visualization for, i.e. who is your audience?

General Audience