# Chapter 3

April 20, 2025

## 0.1 Hands-On Data Preprocessing in Python

Learn how to effectively prepare data for successful data analytics

AUTHOR: Dr. Roy Jafari

### 0.1.1 Chapter 3: Data

```python
import pandas as pd
customer_df = pd.read_excel('Customers Dataset.xlsx')
customer_df
```

```
[2]:      Customer Name                         Store  \
      0        Abu Irvine   Starbuck - Claremont Village
      1   Colleen Melendez  Starbuck - Claremont Village
      2    Lyla-Rose Ruiz   Starbuck - Claremont Village
      3  Riley-Jay Manning  Starbuck - Claremont Village
      4      Ieuan Carroll  Starbuck - Claremont Village
      5   Renesmae Lawson   Starbuck - Claremont Village
      6   Lawrence Medina   Starbuck - Claremont Village
      7      Ben O'Connor   Starbuck - Claremont Village
      8        Adnaan Kim   Starbuck - Claremont Village
      9   Abbigail Dunlap   Starbuck - Claremont Village

         Last week number of visits  Last week Purchase $
      0                           5                 33.43
      1                           4                 11.32
      2                           1                  9.48
      3                           2                 15.50
      4                           4                 17.96
      5                           5                 19.84
      6                           3                 23.21
      7                           1                  6.12
      8                           6                 36.16
      9                           2                  6.88
```
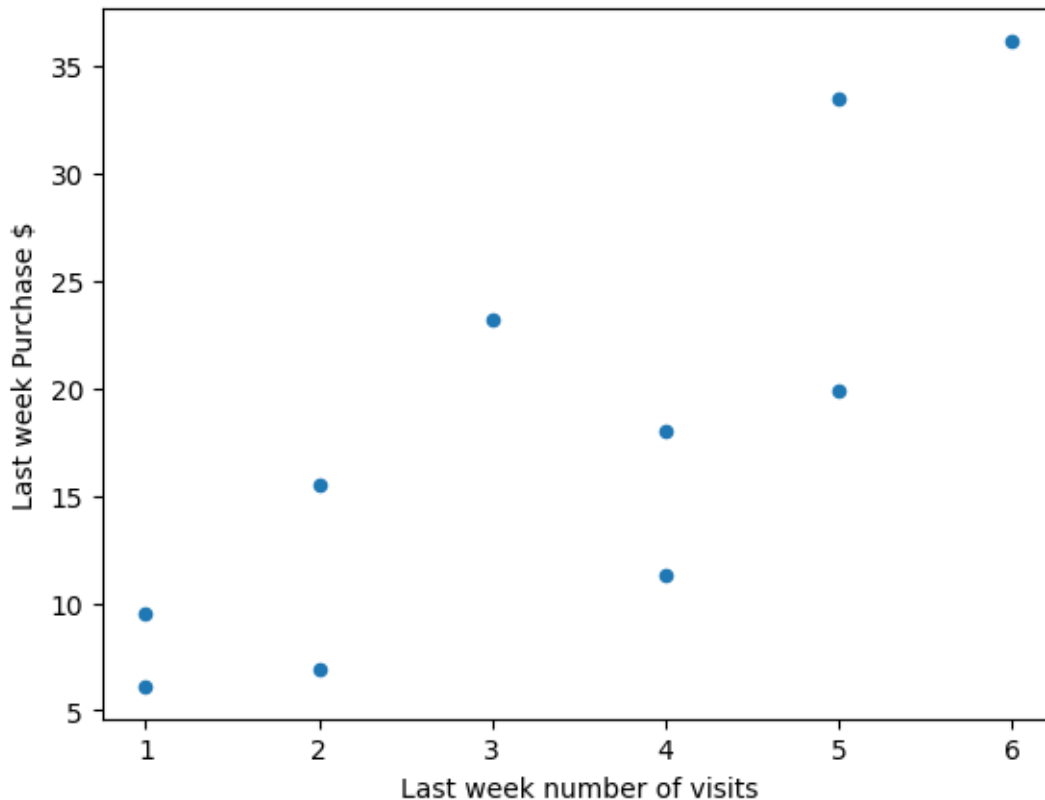
```python
customer_df.iloc[:, 2:].var()
```

```
[15]: Last week number of visits      3.122222
      Last week Purchase $           109.628044
      dtype: float64
```

```
[16]: import matplotlib.pyplot as plt
      customer_df.plot.scatter(x='Last week number of visits',
                               y='Last week Purchase $')
      plt.show()
```



```
[19]: customer_df.iloc[:, 2:].corr()
```

```
[19]:                              Last week number of visits  Last week Purchase $
      Last week number of visits                    1.000000              0.821282
      Last week Purchase $                          0.821282              1.000000
```

```
[20]: customer2_df = pd.read_excel('Customers Dataset 2.xlsx')
      customer2_df
```

```
[20]:      Customer Name                             Store  Last week number of visits  \
      0     Nelson Rivera  Starbuck - Claremont Village                           3
      1     Abbigail Felix  Starbuck - Claremont Village                          1
```
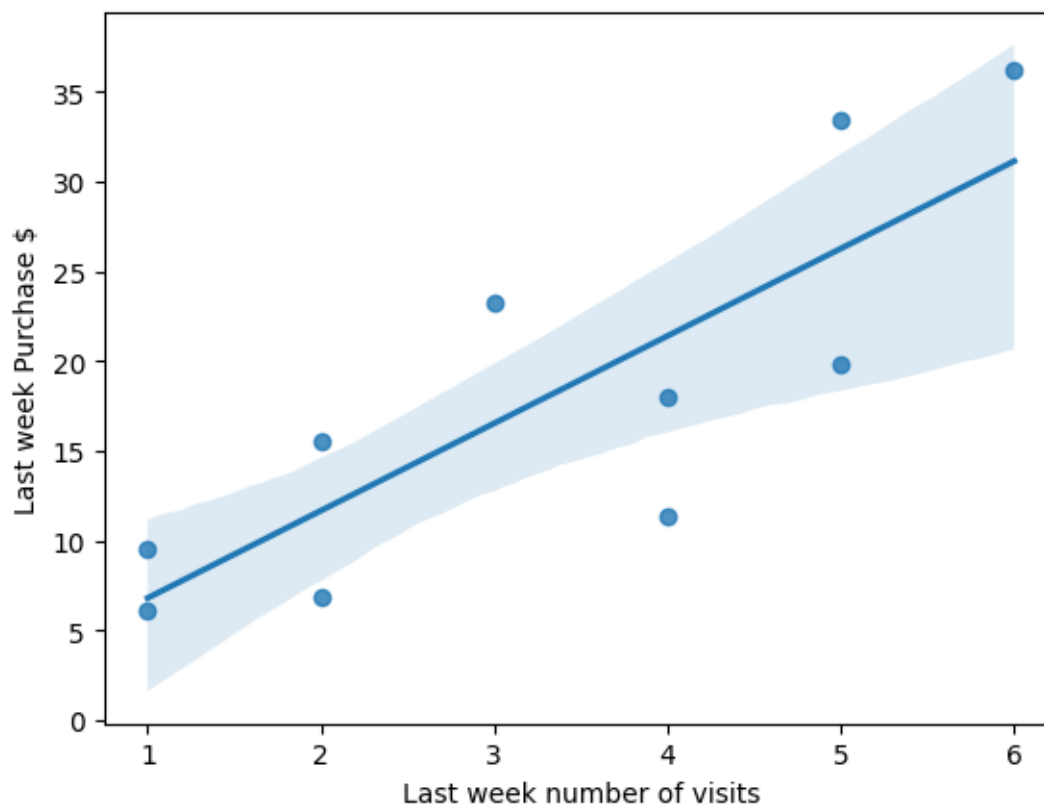
```
2        Kelly North  Starbuck - Claremont Village                2
3        Aneesa Moran  Starbuck - Claremont Village                5
4        Ammara Ritter  Starbuck - Claremont Village               7
5  Elise Valenzuela  Starbuck - Claremont Village                1
6          Jaidan Gay  Starbuck - Claremont Village               4
7  Alejandro Mercer  Starbuck - Claremont Village                3
8  Arisha Whittaker  Starbuck - Claremont Village                5
9      Mehmet Power  Starbuck - Claremont Village                2

   Last week Purchase $
0                  NaN
1                  NaN
2                  NaN
3                  NaN
4                  NaN
5                  NaN
6                  NaN
7                  NaN
8                  NaN
9                  NaN
```

[21]:
```python
import seaborn as sns
sns.regplot(x='Last week number of visits',
            y='Last week Purchase $',data=customer_df)
plt.show()
```

```
[22]: customer2_df['Last week Purchase $'] =  customer2_df[
          'Last week number of visits'].apply(lambda v:1.930 + 4.867 *v)
      customer2_df
```

```
[22]:       Customer Name                        Store  Last week number of visits  \
      0     Nelson Rivera  Starbuck - Claremont Village                           3
      1    Abbigail Felix  Starbuck - Claremont Village                           1
      2       Kelly North  Starbuck - Claremont Village                           2
      3      Aneesa Moran  Starbuck - Claremont Village                           5
      4     Ammara Ritter  Starbuck - Claremont Village                           7
      5  Elise Valenzuela  Starbuck - Claremont Village                           1
      6        Jaidan Gay  Starbuck - Claremont Village                           4
      7  Alejandro Mercer  Starbuck - Claremont Village                           3
      8  Arisha Whittaker  Starbuck - Claremont Village                           5
      9      Mehmet Power  Starbuck - Claremont Village                           2

         Last week Purchase $
      0                16.531
      1                 6.797
      2                11.664
      3                26.265
```

```
4              35.999
5               6.797
6              21.398
7              16.531
8              26.265
9              11.664
```