Seatwork

```
filepath = '/content/nike_shoes_sales.csv'


import pandas as pd
import numpy as np


data = pd.read_csv(filepath)


data
```

- -

Next steps: ◉ View recommended plots

```
#The listed columns of the data set
data.columns
```
```
      Index(['product_name', 'product_id', 'listing_price', 'sale_price', 'discount',
             'brand', 'description', 'rating', 'reviews', 'images'],
            dtype='object')
```

```
#The data type of each column
data.dtypes
```
```
      product_name       object
      product_id         object
      listing_price       int64
      sale_price          int64
      discount            int64
      brand              object
      description        object
      rating            float64
      reviews             int64
      images             object
      dtype: object
```

```
#Total of the shoe model or the product name of nike shoes in the data set
len(data)
```

    643

```
#displays the firts 5 of the datasets
data.head()
```

```
data.head(20)
```

```
#displays the last 5 of the dataset
```

```
#displays the last 5 of the dataset
data.tail()
```

```
data.tail(20)
```

```
#dropping a data column
data.drop(['discount'], axis = 1, inplace =True)
#shows error because the code ran twice, and the datas are already dropped
```

data

```
#dropping multiple data columns
data.drop(['brand','images'], axis = 1, inplace = True)
#shows error because the code ran twice, and the datas are already dropped
```

data

```
#Printing the first 2 and last 2 datas of the dataset
small_data = pd.concat([data.iloc[:2], data.iloc[-2:]])
small_data
```

```
#The mean of the dataset
data.mean()
```

```
<ipython-input-42-1f82bbebc172>:2: FutureWarning: The default value of numeric_only in DataFrame.mean is depi
  data.mean()
listing_price      3875.762053
sale_price        10213.676516
rating                2.734837
reviews               7.181960
dtype: float64
```

```
#The median of the dataset
data.median()
```

```
<ipython-input-46-0c820d1115bf>:2: FutureWarning: The default value of numeric_only in DataFrame.median is de
  data.median()
listing_price         0.0
sale_price         9597.0
rating                3.8
reviews               1.0
dtype: float64
```

```
#The mode of the dataset
data.mode()
```

```
#Overall result of the dataset
data.describe()
```

```
data
```

```
#Adding a new column named Review Checker which shows if the products has a review
data['Review Checker'] = np.where(data['reviews'] == 0, 'No Reviews', 'With Reviews')
data
```

```python
df = pd.DataFrame(data)
basic_data = df[df['product_name', 'sale_price', 'rating']]
basic_data
```

```python
#Prints all the data that has a Review
df = pd.DataFrame(data)
with_reviews = df.loc[df['Review Checker'] == 'With Reviews']
with_reviews
```

```
#Prints all the data that has no Reviews
df = pd.DataFrame(data)
without_reviews = df.loc[df['Review Checker'] == 'No Reviews']
without_reviews
```

```
#Prints the mean of the sale price of the shoe models
np.mean(df['sale_price'])
```

```
10213.676516329704
```

```
#Prints the mean for the listing price
np.mean(df['listing_price'])
```

```
3875.7620528771386
```