ROBOTICS MASTERS

>>> Using Python in e-commerce business

>>> Women in Python Shenzhen

Name: Saulo Meirelles [†]

Date: April 17, 2019

[†]WeChat ID: Saulomini

>>> Learning goals

ROBOTICS MASTERS

1. How Python can help you in your business

>>> Learning goals

ROBOTICS MASTERS

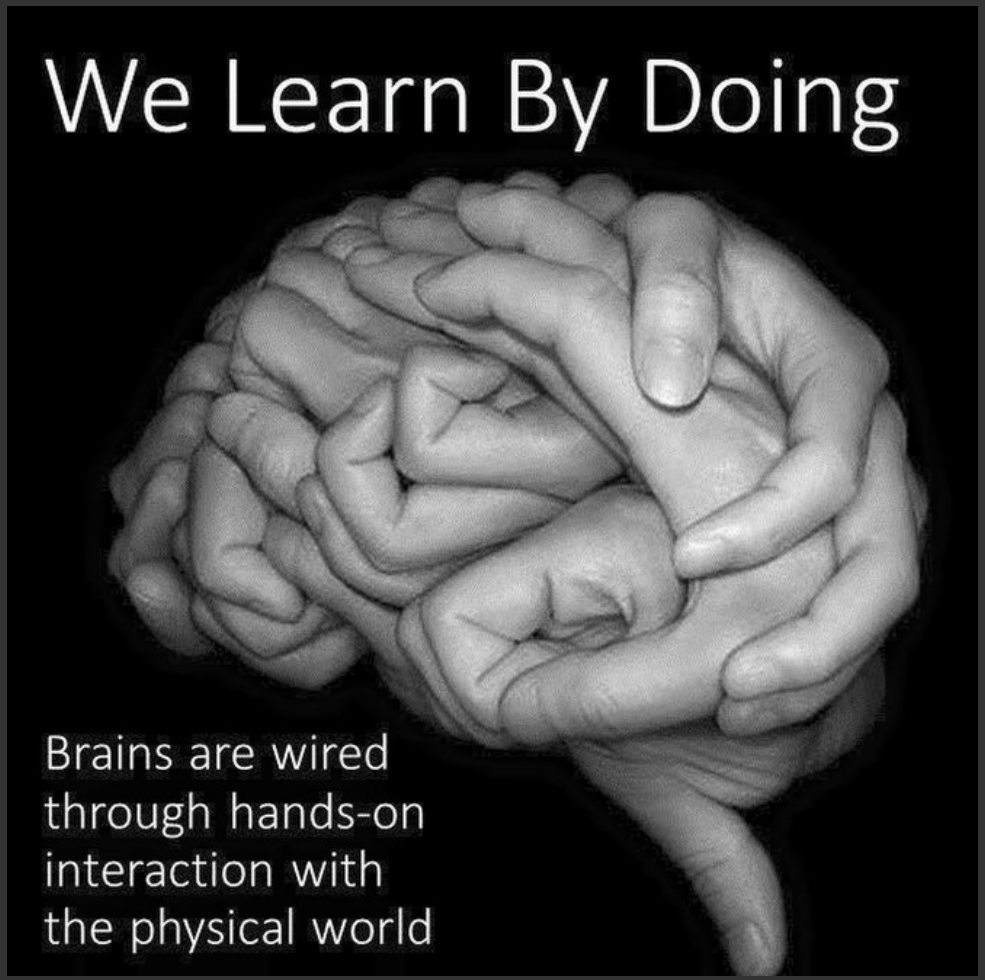
1. How Python can help you in your business

2. Work with Pandas and Scikit-learn

>>> Approach

ROBOTICS MASTERS

Try yourself! ☺



>>> Python in business

ROBOTICS MASTERS

Automation, prediction, organization, etc.

>>> Tonight's goal

ROBOTICS MASTERS

Create a simple Linear Regression Model to analyze the main revenue sources of an e-commerce business.
We are going to need Python's Pandas and Scikit-learn libraries.

Packages to install

```
conda install scikit-learn
conda install pandas
```

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer's email id)
2. Address (Customer's home address)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer's email id)
2. Address (Customer's home address)
3. Avatar (member profile avatar)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)
2. Address (Customer' s home address)
3. Avatar (member profile avatar)
4. Average session length (Minutes spent by customer on average for each in-store session)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)
2. Address (Customer' s home address)
3. Avatar (member profile avatar)
4. Average session length (Minutes spent by customer on average for each in-store session)
5. Time on App (Minutes spent by customer on the app)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)
2. Address (Customer' s home address)
3. Avatar (member profile avatar)
4. Average session length (Minutes spent by customer on average for each in-store session)
5. Time on App (Minutes spent by customer on the app)
6. Time on Website (Minutes spent by customer on the website)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)
2. Address (Customer' s home address)
3. Avatar (member profile avatar)
4. Average session length (Minutes spent by customer on average for each in-store session)
5. Time on App (Minutes spent by customer on the app)
6. Time on Website (Minutes spent by customer on the website)
7. Length of Membership (Years the customer has been a member)

[1] Save the file in the same folder where your Python Notebook is.

>>> Our dataset

ROBOTICS MASTERS

Please, download our dataset from our repo: https://github.com/saulomeirelles/WomenInPythonShenzhen[1]

1. Email (Customer' s email id)
2. Address (Customer' s home address)
3. Avatar (member profile avatar)
4. Average session length (Minutes spent by customer on average for each in-store session)
5. Time on App (Minutes spent by customer on the app)
6. Time on Website (Minutes spent by customer on the website)
7. Length of Membership (Years the customer has been a member)
8. Yearly Amount Spent (Money spent yearly by customer)

[1] Save the file in the same folder where your Python Notebook is.

ROBOTICS MASTERS

Remember the first thing to do is to import the libraries (packages) we are going to work with.

Importing libraries

```
# pandas is a Python package providing powerful data structures
import pandas as pd
# NumPy is a package for scientific computing with Python
import numpy as np
# Matplotlib is a Python plotting library for quality figures
import matplotlib.pyplot as plt

# Scikit provides tools for Machine Learning and data analysis
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

ROBOTICS MASTERS

Let's making sense of our dataset.

Basic commands

```
df_ecom = pd.read_csv('EcommerceCustomers.csv') # reading the
    data set (csv file)
df_ecom.info() # prints information about a DataFrame
df_ecom.describe() # Generate descriptive statistics that
    summarize the central tendency, dispersion and shape of a
    dataset's distribution
```

ROBOTICS MASTERS

Let's now check the customers loyalty by analyzing the length of the membership.
Is the length of the membership correlated with the yearly amount spent?

Basic commands

```
# more than 6 year as a member
df_ecom.loc[df_ecom['Length of Membership'] > 6].plot.bar(x='
    Avatar',y='Length of Membership')
# less than 1 year as a member
df_ecom.loc[df_ecom['Length of Membership'] < 1].plot.bar(x='
    Avatar',y='Length of Membership')
# relation between amount spent and length of membership
df_ecom.plot.scatter(x='Yearly Amount Spent',y='Length of
    Membership')
```

ROBOTICS MASTERS

Remember our goal is to analyze the 'Yearly amount spent' (y) for each customer which is as function of 'Time on App' (x_1), 'Time on Website' (x_2), 'Length of Membership' (x_3).

Define y(x1,x2,x3)

```
# values of x's
x = df_ecom[['Time on App', 'Time on Website', 'Length of
    Membership']]
# values of y
y = df_ecom[['Yearly Amount Spent']]
```

ROBOTICS MASTERS

You may remember from last class that we had 70,000 images of which 10,000 were used for training?
Here, we will use 30% of the dataset for testing and the remaining 70% for training. There is a command for that in Scikit:

Split the dataset

```
X_train, X_test, y_train, y_test = train_test_split(x, y,
    test_size=0.3, random_state=101)
# check what is in X_train
X_train.head()
```

>>> Working with Scikit-learn

ROBOTICS MASTERS

Let's build, train, and test our model!

Ordinary least squares Linear Regression

```
lm = LinearRegression()
lm.fit(X_train, y_train)
predictions = lm.predict(X_test)

plt.scatter(y_test, predictions)
plt.xlabel('Y Test')
plt.ylabel('Predicted Y')
```

ROBOTICS MASTERS

Let's check the results!

Coefficients

```
coeffs = pd.DataFrame(data=lm.coef_.transpose(), index=X.columns,
    columns=['Coefficient'])
```

Highly correlated features have the highest coefficients in the model, while features uncorrelated should have coefficient values close to zero.
The values of our coefficients tell us that one minute on the app corresponds to $36.75 in revenue, whereas one minute on the website corresponds to just $1.07 in revenue.

ROBOTICS MASTERS

Try yourself!

Predictions

```
lm.predict(x_new_values)
```

Play around with the predictions and see how they can be applied in your business.

References:

https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.h

https://medium.com/tensorist/making-e-commerce-business-decisions-usi