

# Python for Data Analytic

Mohd Saufy Rohmad  
Smart Manufacturing Research Institute  
UiTM



اَوْنِيُوْرَسِيْتِي تِيْكُوْلُوْكِى مَارَا  
UNIVERSITI  
TEKNOLOGI  
MARA

Institut  
Penyelidikan  
Pembuatan Pintar

**SMRI**

■ Smart  
■ ■ Manufacturing  
■ ■ ■ Research Institute

# Installing Pandas on Windos

- <https://data-flair.training/blogs/install-pandas-on-windows/>
- Install pandas on your pc by following above website
- Please follow and test your pandas installation.

# What is Big Data Analytic

- Python is powerful language for data analysis.
-

# Pandas(1)

- Pandas provide rich data structure and function designed to make working with structured data fast, easy and expressive.
- The primary object in pandas that will be used in this book is the *Dataframe*, a two dimensional tabular, column-oriented data structure with both row and column labels.

# Pandas (2)

```
>>> frame
```

```
total_bill tip sex smoker day time size
```

```
1 16.99 1.01 Female No Sun Dinner 2
```

```
2 10.34 1.66 Male No Sun Dinner 3
```

```
3 21.01 3.5 Male No Sun Dinner 3
```

```
4 23.68 3.31 Male No Sun Dinner 2
```

```
5 24.59 3.61 Female No Sun Dinner 4
```

```
6 25.29 4.71 Male No Sun Dinner 4
```

```
7 8.77 2 Male No Sun Dinner 2
```

```
8 26.88 3.12 Male No Sun Dinner 4
```

```
9 15.04 1.96 Male No Sun Dinner 2
```

```
10 14.78 3.23 Male No Sun Dinner 2
```

# Pandas(3)

- Pandas combine the high performance array-computing features of NumPy with the flexible data manipulating capabilities of spreadsheets and relational databases (such as SQL).
- It provide sophisticated indexing function to make it easy to reshape,slice and dice,perform aggregations, and select subsets of data.

# Data Analysis Process

- ***Interacting with outside world*** - Reading and writing with a variety of file formats and databases.
- ***Preparation*** – cleaning, munging, combining, transforming the data
- ***Transformation*** – Applying mathematical and statistical operations to groups of dataset
- ***Modeling and Computation*** – Connecting your data to statistical model, machine learning algorithms.
- ***Presentation*** – Creating interactive or static graphical visualizations or textual summaries.

# Let's begin with sample dataset

- MovieLens 1M Data Set
  - The data provide movie ratings, movie metadata (genres and year), and demographic data about the users (age, zip code, gender, and occupation).
  - The MovieLens 1M data set contains 1 million ratings collected from 6000 users on 4000 movies.
  - It's spread across 3 tables: ratings, user information, and movie information.
  - After extracting the data from the zip file, each table can be loaded into a pandas DataFrame object using ***pandas.read\_table***



# Read Data

```
import pandas as pd

unames = ['user_id', 'gender', 'age', 'occupation', 'zip']
users = pd.read_table('ml-1m/users.dat', sep='::', header=None,
names=unames)

rnames = ['user_id', 'movie_id', 'rating', 'timestamp']
ratings = pd.read_table('ml-1m/ratings.dat', sep='::', header=None,
names=rnames)

mnames = ['movie_id', 'title', 'genres']
movies = pd.read_table('ml-1m/movies.dat', sep='::', header=None,
names=mnames)
```

# Print data

```
>> users
```

```
>> ratings
```

```
>> movies
```

# Print data types

```
>> ratings.dtypes
```

# Merge Data

- `data=pd.merge(pd.merge(ratings,users),  
movies)`

# Print data on row

```
>>> data[0:1]
```

user_id	movie_id	rating	timestamp	gender	age	occupation	zip	title
0	1	1193	5	978300760	F			
1	10	48067	One Flew Over the Cuckoo's Nest (1975)	Drama				

# Get Mean Rating

- `>>mean_ratings=data.pivot_table('rating',index='title',columns='gender',aggfunc='mean')`
- `>> mean_ratings`

# Rating by Tittle

```
>>ratings_by_title = data.groupby('title').size()
```

```
>>ratings_by_title
```

# Load MySQL table to dataframe

```
import mysql.connector as sql
import pandas as pd

db_connection = sql.connect(host='localhost', database='epo666_IoT',
                             user='root', password='saufy1982')
db_cursor = db_connection.cursor()
db_cursor.execute('SELECT * FROM touch_sensor')
table_rows = db_cursor.fetchall()
df = pd.DataFrame(table_rows)
print(df)
```



# Data Frame Processing

- Data Loaded into dataframe format can be process like the example before
- This work present how to link the data collected from mysql to panda data analytic library.
- The process can be further enhance with machine learning algorithms.

# Conclusion

-