|  |  |
| --- | --- |
| **Practicum Case** |  |
| COMP6579  Big Data Processing |
| **Computer Science** | **E201-COMP6579-DD01-11** |
| ***Valid on*** *Even Semester Year 2019/2020* | **Revision 00** |

## Learning Outcomes

* Understand Big Data Analytics and Visualizations

## Topic

* Session 11 – Data Visualization

## Subtopics

* Data Visualization in Python

## Soal

*Case*

**Ramen Shop**

**Ramen Shop** is a store located in Jakarta. Since the sales are going high, they need to improve the store sales more quickly. To do that they intended to do some **analysis** from **different** **kinds** of data they have.

From the sales business process, the data can be analyzed to gain sales insight. The data is stored in **Comma-Separated** **Values (CSV)** file and the data schema is drawn using **Entity Relationship Diagram (ERD)** below:

*A screenshot of a social media post

Description automatically generated*

**Figure 1. Ramen Shop ERD**

You were given the task to gain some insight from the sales data. Below is the task you must do:

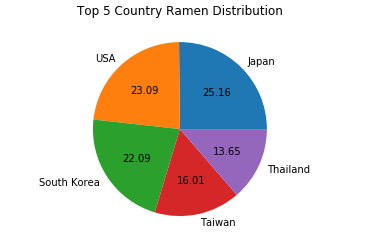
1. **Load Data from CSV to Spark**

Using **SparkSession**, **read** the following files (“MsBrand.csv”, “MsCountry.csv”, “MsCustomer.csv”, “MsRamen.csv”, “MsRamenStyle.csv”, “TransactionHeader.csv”, “TransactionDetail.csv”).

1. **Query Analysis and Visualization**

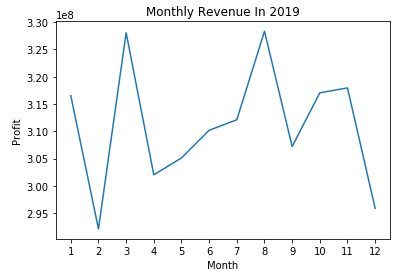
You are asked to gain some sales insight about the data. Below are some statements you need to answer. Use **SparkSQL** to answer the question and **pyplot** package to **visualize** the answer.

1. Show **top 5 country** which **distributed** **the most number of ramen** using **pie plot**. Don’t forget to add **title**, **labels,** and **percentage** for the plot.



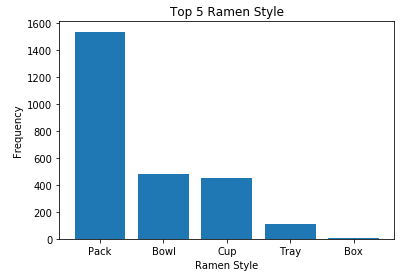
**Figure 2. Top 5 Country Ramen Distribution Figure**

1. Show the **amount of revenue** for **each month** within **year 2019** using **line plot**. Don’t forget to add **title**, **x-label**, and **y-label** for the plot.



**Figure 3. Monthly Revenue in 2019 Figure**

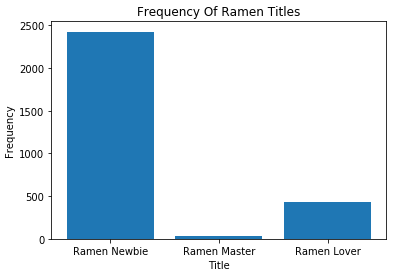
1. Show **top 5 ramen style** that is **used in making the ramen** using **bar plot**. Don’t forget to add **title**, **x-label**, and **y-label** for the plot.



**Figure 4. Top 5 Ramen Style Figure**

1. Show the **frequency of customer** who earned “**Ramen Master**”, “**Ramen Lover**”, and “**Ramen Newbie**” **title** using **bar plot**. Don’t forget to add **title**, **x-label**, and **y-label** for the plot. The **title** is achieved based on the following condition:

|  |  |
| --- | --- |
| **Ramen Ordered** | **Title** |
| > 125 | Ramen Master |
| 76 - 125 | Ramen Lover |
| 1 - 75 | Ramen Newbie |



**Figure 5. Frequency of Ramen Titles Figure**

**Please ask your teaching assistant if there are any related questions.**