Business Analytics Practicum (MGT 4803)

Zhaohu (Jonathan) Fan

Information Technology Management

Scheller College of Business

Georgia Institute of Technology

January 25, 2024

Sparck Technology

- **The Boxsters (Team #2)**: Sparck Technology (project focused on event analytics with detailed visualizations for each machine at a micro level)
- **Scheller Sparck Squadron (Team #4)**: Sparck Technology (project focused on analyzing and presenting machine health and performance metrics at a macro level)

Address this business case/challenge

- **Step 1: Defining the project scope and objectives** - It's crucial to specify which machines and events we're focusing on. For instance, are we examining packaging machines such as the CVP Impack and CVP Everest?
- **Step 2: Data collection and preparation** - Identifying available data sources and understanding how to access them is key. Furthermore, we need to establish the steps required for data cleaning and preprocessing.

What is the average height of US presidents?

- **Pandas is a library for** data manipulation **and data analysis.**
- **Numpy** is an **array manipulation** library, used for linear algebra, Fourier transform, and random number capabilities.
- Matplotlib is a library which generates **figures** and provides graphical user interface toolkit.

What is the average height of US presidents?

- Please click on the link provided below
  - In-Class Exercise 1: NumPy, Pandas, Matplotlib

Reading a CSV file with Pandas

Python code

```
>>>data_df = pd.read_csv('XXXXXXX.csv', sep=';')
```

- `pd.read_csv()`: is a function provided by pandas to read a comma-separated values (CSV) file into a DataFrame.
  - The `sep=';'` (separators): indicates that the file uses semicolons (;) instead of the more common commas to separate its columns.

Reading a .rda file with Pandas

Python code

- installing `pyreadr` (exclamation mark !pip install pyreadr)
  - `pyreadr.read_r` to read the file into a pandas DataFrame.

Reading data sets using R

- Read the CSV File: Use the `read.csv()` function to load the CSV file into a data frame.
  - View the Data: Use functions like `head()` to view the first few rows of the data frame.
  - Use the `load()` function to read your .rda file.

R code

```
# Load the necessary library
library(readr)
# Read the CSV file
df ← read.csv("XXX.csv")
# Display the first few rows of the data frame
head(df)

load("myData.rda")
print(myData)
```