

# Business Analytics Practicum (MGT 4803)

Zhaohu (Jonathan) Fan

Information Technology Management

Scheller College of Business

Georgia Institute of Technology

February 13, 2024

- Company site visit for undergrad business analytics practicum project
  - **When:** March 1st (Friday)
    - 10:00 AM - 1:00 PM (with a maximum duration of 3 hours)
    - Catering by Sparck during the visit
  - **Where:** Sparck at the location in Marietta, GA
  - **How to proceed?** Signing the GT waiver (which can be found in Module 6, Week 5)

- **Gain insights** into CVP systems' functionality, bridging your work with field machinery
- **Engage in discussions** on KPIs and thresholds to enhance teamwork and productivity
- Seize the chance to connect with the product engineers and service manager

## Challenge for Sparck Technology Teams: Efficient handling of large datasets

- **Objective:** Develop a method to read hundreds of CSV files simultaneously.
- **Key Question:** How can we smartly read hundreds of CSV files all at once, instead of one by one?

- **Key Question: How can we smartly read hundreds of CSV files all at once, instead of one by one?**
- Using `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.

# Regular Pattern(regex pattern)?

- `^`: Asserts the start of a string, ensuring matches begin from the start of the filename.
  - `3dretractions_2023-10-`: Matches files starting explicitly with "3dretractions\_2023-10-", indicating files related to 3D retractions from October 2023.
  - `.*`: The dot `.` matches any character (except newline), and `*` allows for zero or more occurrences of any character, accommodating various characters following the initial prefix.
  - `\\.csv`: Matches the file extension ".csv". The backslash `\\` is used to escape the dot `.` in the regex pattern, as the dot is a special character in regex. The backslash itself must be escaped in R strings, hence `\\.`
  - `$`: Asserts the end of a string, ensuring the pattern matches only if the filename ends with ".csv".

## R code

```
# Pattern to match files of interest
```

```
pattern ← "^3dretractions_2023-10-.*\\.csv$"
```

```
# List all files in the directory that match the pattern
```

```
file_paths ← list.files(pattern = pattern)
```

# R Demonstration