Business Analytics Practicum (MGT 4803)

Zhaohu (Jonathan) Fan

Information Technology Management

Scheller College of Business

Georgia Institute of Technology

February 29, 2024

Logistical updates: Sparck Technology (Demo Day)

- 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

First presentation (Data Exploration and Visualization)

- **When**: 5:00PM – 6:15PM, March 7th (Thursday)

  - **Where**: Scheller College of Business 203

  - **More details?** Please refer to the Team Mid-Point Presentation

    - **Canvas->Modules-> Assignments ->Team Mid-Point Presentation #1: (Data Exploration and Visualization)**
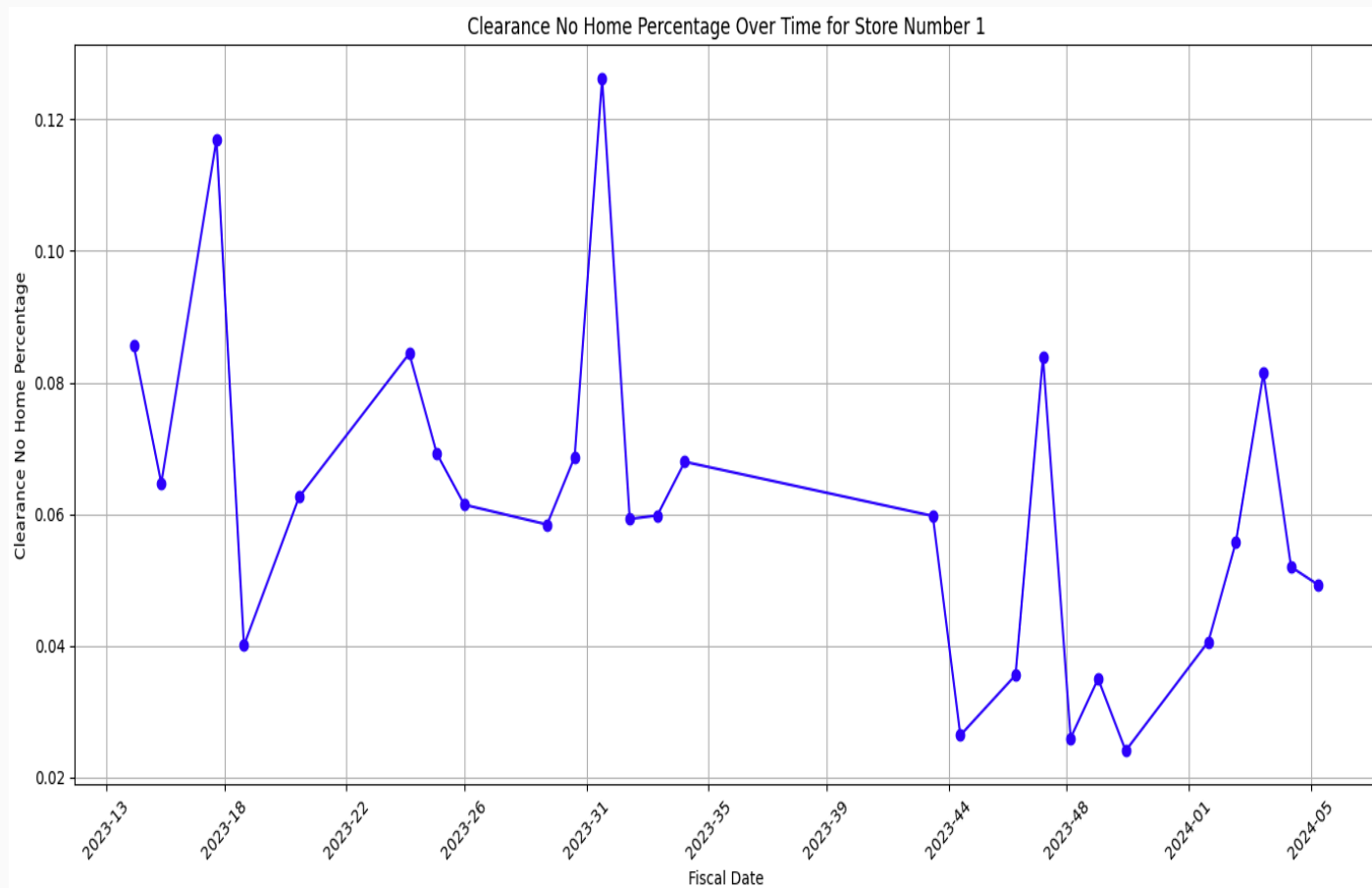
Time series data visualization

- How does the `CLEARANCE_NO_HOME_PCT` trend change throughout the year?

  - Let's take a look at the fiscal weeks.

  - **On our x-axis, we have fiscal dates from the 13th week of 2023 to the 5th week of 2024. What significant events in this period could affect our percentages?**
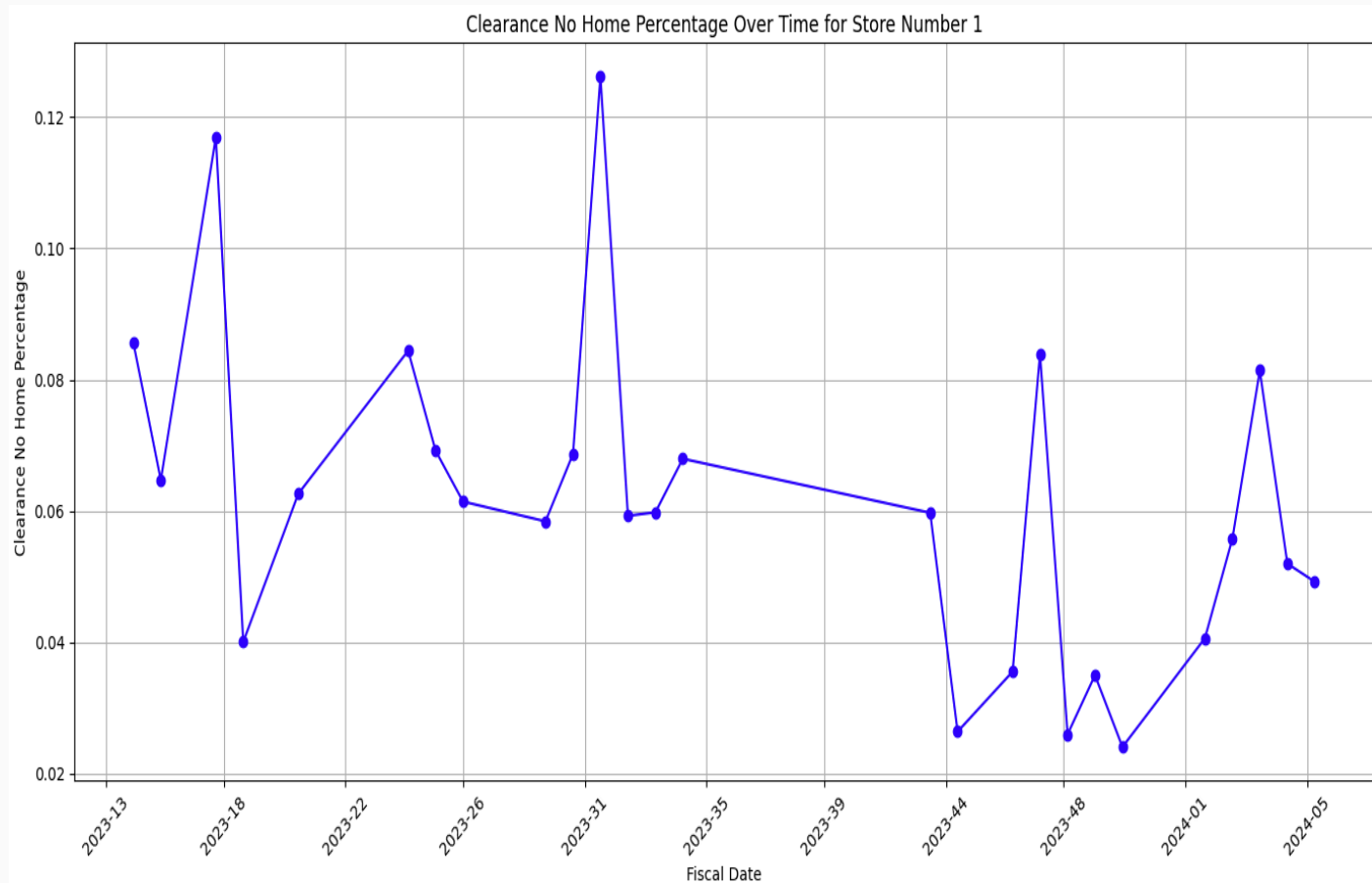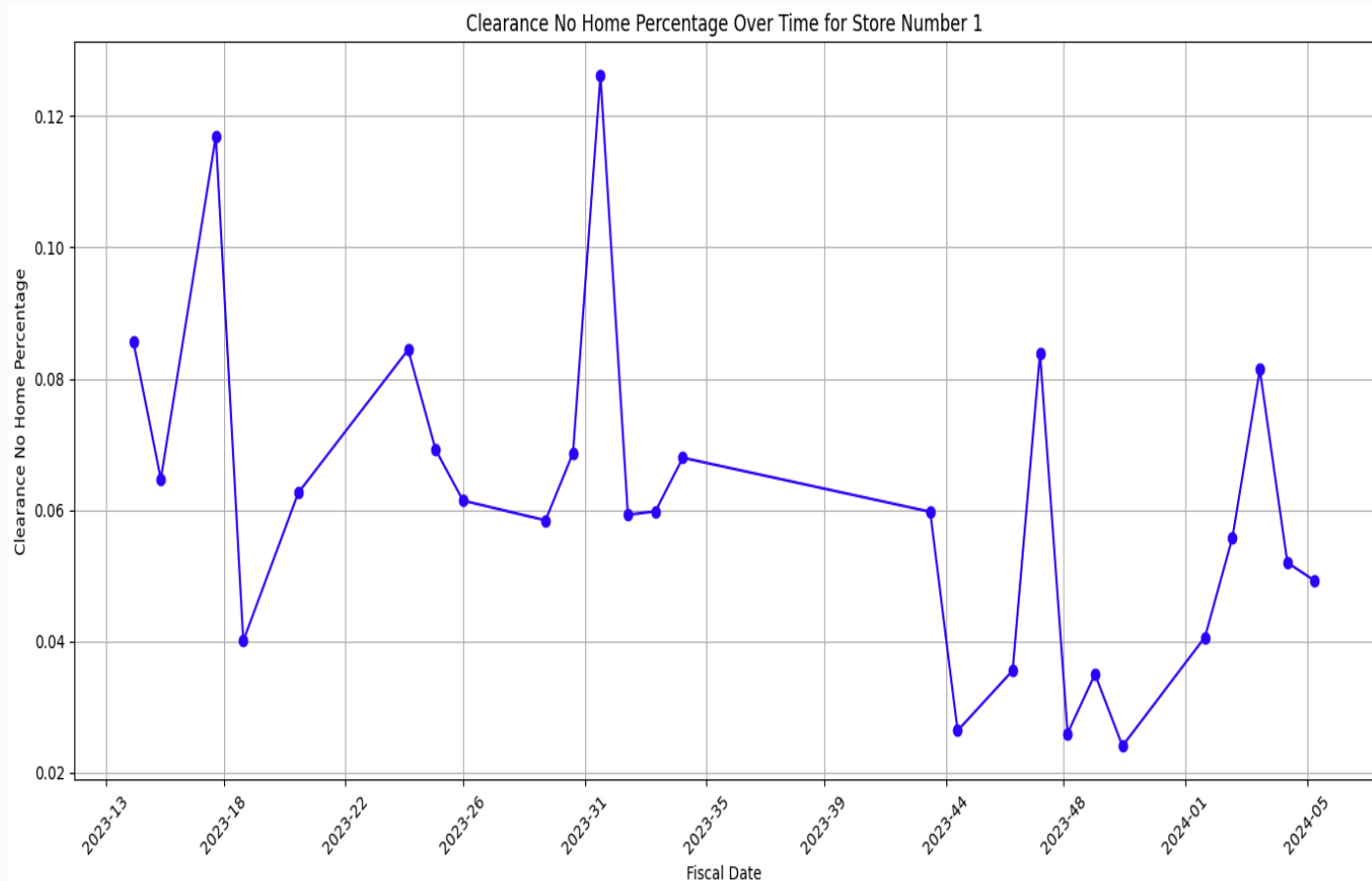
Time series data visualization

- How does the `CLEARANCE_NO_HOME_PCT` trend change throughout the year?

  - Let's take a look at the fiscal weeks.

  - **Can anyone guess which times of year we might see peaks or drops?**


Clearance No Home Percentage Over Time for Store Number 1

Time series data visualization

- **Looking at the trend of** `'Clearance No Home Percentage'` **, what fluctuations do you anticipate during fiscal weeks?**

Clearance No Home Percentage Over Time for Store Number 1

Time series data visualization

- When comparing total tasks assigned to those completed, what story do you think the data will tell us about our performance over time?

  - **What insights do you expect we'll gain for Store Number 1?**

Clearance No Home Percentage Over Time for Store Number 1
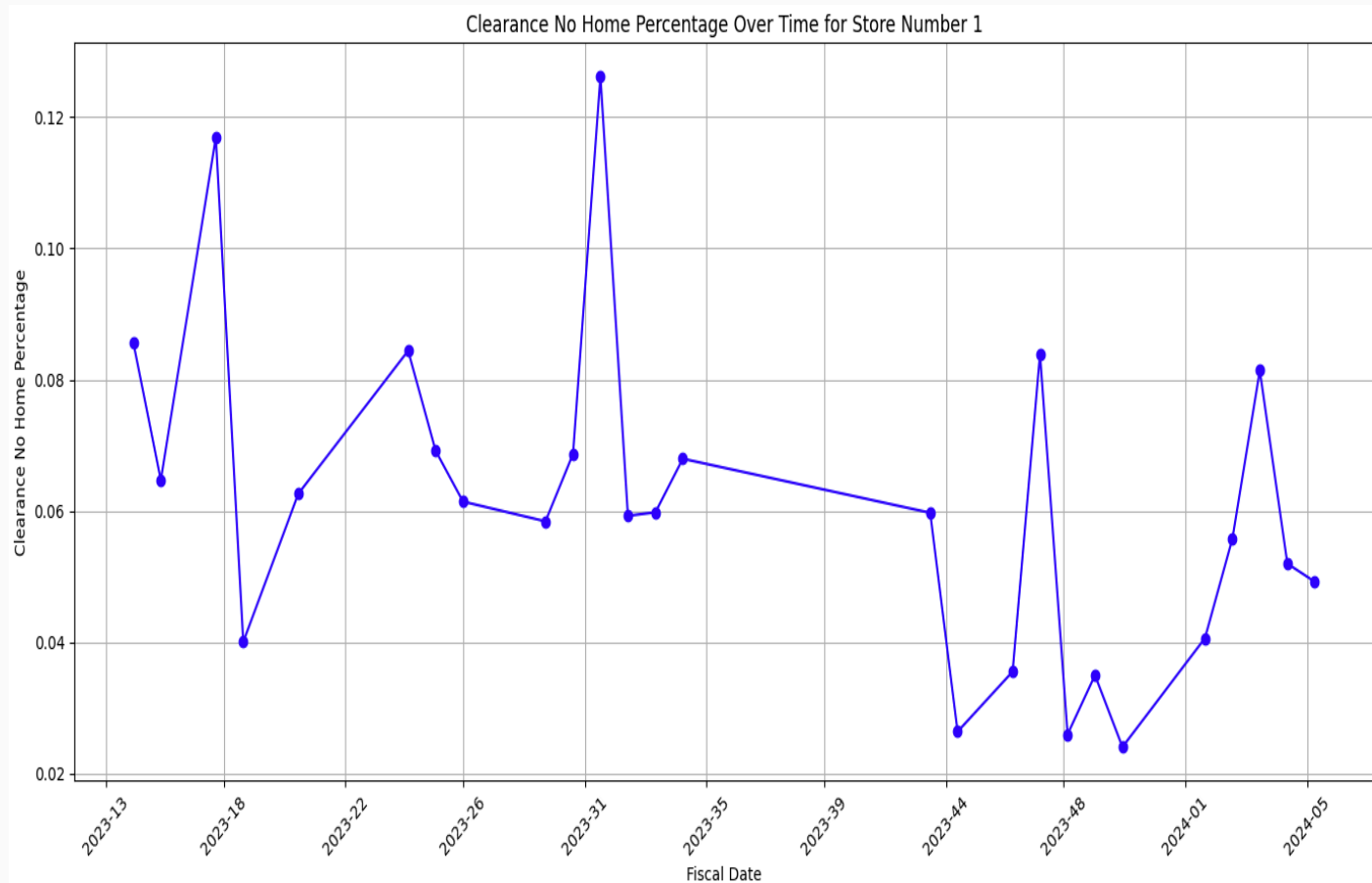
Appendix

Why is the 'WEEK' column not sorted in ascending order, and how can we address this issue, especially if the 'WEEK' column is sorted as strings rather than numbers?

Our proposal

- Convert 'YEAR' and 'WEEK' to Strings

  - Extract Numerical Values from 'YEAR' and 'WEEK'

  - Sort the DataFrame

  - Drop Temporary Columns

  - Display the Sorted DataFrame

Convert 'YEAR' and 'WEEK' to Strings

- The code converts the columns `YEAR` and `WEEK` in the DataFrame `df_joined3` to string data types.

  - This is done using the .astype(str) method. This step is necessary if these columns are not already in string format, especially if subsequent operations rely on them being strings.

Python code

```python
# Convert 'YEAR' and 'WEEK' to string if they're not already
df_joined3['YEAR'] = df_joined3['YEAR'].astype(str)
df_joined3['WEEK'] = df_joined3['WEEK'].astype(str)
```

- The DataFrame df_joined3 is then sorted based on three columns: `'STR_NBR'`, `'YEAR_NUMBER'`, and `'WEEK_NUMBER'`. The `sort_values(by=[ ... ])` method is used for this purpose.

  - This means the DataFrame will first be sorted by `'STR_NBR'`, within that by `'YEAR_NUMBER'`, and finally within each year by `'WEEK_NUMBER'`.

Python code

```python
# Extract numerical part of 'YEAR' and 'WEEK' for sorting
df_joined3['YEAR_NUMBER'] = df_joined3['YEAR'].str.extract('(\d+)').astype(int)
df_joined3['WEEK_NUMBER'] = df_joined3['WEEK'].str.extract('(\d+)').astype(int)
```

Sort the DataFrame:

`

- The DataFrame df_joined3 is then sorted based on three columns: `'STR_NBR'`, `'YEAR_NUMBER'`, and `'WEEK_NUMBER'`.

- `The sort_values(by=[ ... ])` method is used for this purpose.

Python code

```python
df_sorted = df_joined3.sort_values(by=['STR_NBR', 'YEAR_NUMBER', 'WEEK_NUMBER'])
```

Python (Google Colab) Demonstration

- Please click on the link provided below

  - Home Depot Project