William and Mary's Union Central Automation Project

Andrew Choi
Data Science
College of William and Mary
Williamsburg, Virginia
ajchoi02@wm.edu

David Cho
Data Science, Math
College of William and Mary
Williamsburg, Virginia
dhcho@wm.edu

Carlo Mehegan
Computer Science, Data Science
College of William and Mary
Williamsburg, Virginia
cjmehegan@wm.edu

I. PROJECT DESCRIPTION

A. Project Statement

Union Central, located on the first floor of the Sadler Center, is a dedicated place at the College of William and Mary that offers various types of games, consoles, and fun activities. This is an important place for students as they are able to take time off from their stressful work and relax by themselves or with friends by playing some games and other sort of activities. In previous work with Union Central employees, a project was done where occupancy and managing equipment rentals were logged to be turned into usable data. The start of the data collection during the shifts included information about the use of different game consoles, rentals of various items, and overall occupancy trends. While this project has addressed the start of incorporating data driven analysis for Union Central, such valuable data still remains underutilized.

This project aims to build beyond the scope of the previous project by automating the process of data analysis and visualization for Union Central's website. By creating specific Python scripts and utilizing Google Sheet's API, we plan on building a pipeline that pulls new data, and after cleaning it, returning it back, ideally in real time, to the website for the viewer to see. Our goal with this project is to be able to showcase the ways automation can bring forth progress in a data-driven subject.

B. Project Delivery

The main deliverable will be an interactive visualization dashboard. This visualization is a rough floor plan representing the different areas of Union Central. Users click on different areas to see updated statistics about that section. For example, clicking on the game console will show data on their use. Clicking on the game console area will show data on how the consoles compare to each other. This pop-up information is in the form of pie charts, bar graphs, or other small visualizations along with descriptions. These visualizations will update either at the press of a button or automatically to incorporate new data.

We aim to use the data to create automated visualizations for us. This important task will help to not only provide strong and usable visuals for the audience to see, but also act as a great resource for the Union Central team as this will offer ready to use visualizations that either employees or professional staff can use freely in company-wide meetings each semester.

II. DATA DESCRIPTION

A. Dataset

The dataset currently in use comes from Union Central data collected from the Fall 2023 semester. As we work to automate the dashboard, we will incorporate data from Spring 2024. Once we have a general workflow, we will incorporate data from Fall 2024 as the semester goes on to test live-update capability.

This data is collected by Union Central employees where all the data is saved through Google Sheets. The key features in the rental logs include student names and IDs. Because the data is sensitive with student names, we will anonymize this data by removing the names entirely and giving each student their own identifier (starting with 1, 2, etc.) in order to maintain the relationship between renter and rentals while removing personal identifiers. With the data being taken of William and Mary students, the dataset used in the project is not able to be shared for public use and replication of this project.

B. Data Features

The Union Central dataset is separated into multiple tables. Here is an excerpt from the occupancy data. The key features here are the day, date, headcount, and time of record. Time between occupancy recordings varies, from ten minutes to an hour. Occupancy is typically taken every 20-30 minutes.

Day	Date	Headcount	Time of Record
Friday	8/25/2023	3	7:30
Friday	8/25/2023	0	7:50
Friday	8/25/2023	2	8:14
Friday	8/25/2023	5	8:39
Friday	8/25/2023	14	9:11
Friday	8/25/2023	9	9:25
Friday	8/25/2023	13	9:33
Friday	8/25/2023	12	9:52
Friday	8/25/2023	3	7:30

Fig. 1. Sample of 2,441 Occupancy Records

Pool, Shuffleboard, Air Hockey, and Foosball data is collected in another table. The key features here are the date, game, time in, time out, and the table number.

Date	Name	ID	Game	In	Out	Table #
8/25			Pool	12:14	12:25	1
			Foosball	12:14	12:30	
			Pool	12:41	12:55	3
			Pool	12:58	1:07	2
			Pool	1:02	1:36	3
			Pool	1:09	1:20	3
			Pool	1:12	1:18	1

Fig. 2. Sample of 3,598 rental records for table games

Here is an excerpt, with names and IDs removed. We will use a python script to systematically remove the names and IDs, and replace them with unique identifiers. The date in the first column is only recorded for the first entry of each day. We will need to take this into account when parsing the data. The "Table #" column is only used if the game is Pool. The reason we do this is because there are multiple pool tables, but only one shuffleboard, air hockey, and foosball table.

Video game data is collected in another table. Names and IDs have been removed. The key features here include the date, type of console, type of game, number of controllers used, as well as the time in and time out.

Date	Name	ID	Console	Game	Controllers	In	Out
9/2			Xbox	Minecraft	1	12:23	1:15
			Wii	Mario Party 8	4	1:37	2:00
			Wii	Mario Kart	2	2:55	5:02
			Xbox	Minecraft	2	3:45	5:00
			Wii	Splatoon	1	5:15	5:45
			Xbox	Forza Horizon 5	1	7:02	9:44

Fig. 3. Sample of 448 rental records for video games

Board game data is collected in another table. The key features here are the date, type of game, the time in and time out, and finally the notes. Names and IDs have been removed. The "Game" column here is a drop down, which only contains the nine most popular board games.

Date	Name	ID	Game	In	Out	Notes
9/9			Other (specify in Notes)	7:14	8:08	Chess
			Other (specify in Notes)	9:12	9:31	Set
9/10			Other (specify in Notes)	6:40	7:45	Taboo
9/15			Deck of Cards	6:31	6:34	
			Deck of Cards	7:53	8:08	
9/19			Uno	7:18	7:55	

Fig. 4. Sample of 113 rental records for board games

III. PROJECT IMPLEMENTATION PLAN

The current standing for the Union Central project is that we have a strong and usable dataset revolving around the types of games used, occupancy, as well as other rental records. We also have a Github page that portrays an outline of Union Central. The website page is interactive where if you click on an element from the map visualizations appear for the particular game or table you selected. These visualizations describe the data based on what you clicked, but also highlight some comparisons and other forms of analysis that may be interesting to look at.

In addition, the automated script for parsing the UC data is almost complete. We currently have multiple small scripts that currently clean and properly process a specific column of the UC dataset. The goal is to be able to combine all of the scripts into one where we simply have one or a few lines of code that will return a ready to go dataset that will be used for analysis as well as visualizations. To add on, we also have an automated script ready that is able to take the data (using dummy data for now) and convert it from Google Sheets and turn it into a usable csv. This will allow for easier, real time updates to the website for further data analysis. One audience we are catering to is the professional staff member who creates a yearly report using UC's data. Our goal is to automate parts of that process and have the website serve as a supplement or replacement to that report.

Once we have the updated data, we plan on updating the visualizations automatically where specific visuals will be created based on the criteria we specify. This then will be updated onto the webpage for those who want to see the current live version of the Union Central data. If time permits, we are also thinking to have a script that will be able to regularly update the live website, for example every morning, to include the data that was added the day before. This would allow for the website to show almost real-time data and allow the audience to see up to date material on Union Central.

IV. GROUP WORK

Andrew Choi: Using Google Sheets API to automatically take the Google Sheets data and convert that into a usable CSV data file. Adding and updating the website data using the CSV in real time.

David Cho: Creating automatic visualizations based on the updated data. Making sure specific criteria can be specific to create different kinds of images (like png or html)

Carlo Mehegan: Creating the automatic script that will clean and parse through the updated data. Making sure to account for all forms of issues that may come with the Union Central Data

Everyone: Developing functions for analyzing the different datasets into visualizations. Each person will focus on a different subsection of the data and add their sections to the main deliverable and this project proposal.

REFERENCES

[1] https://tinyurl.com/ucdata2024