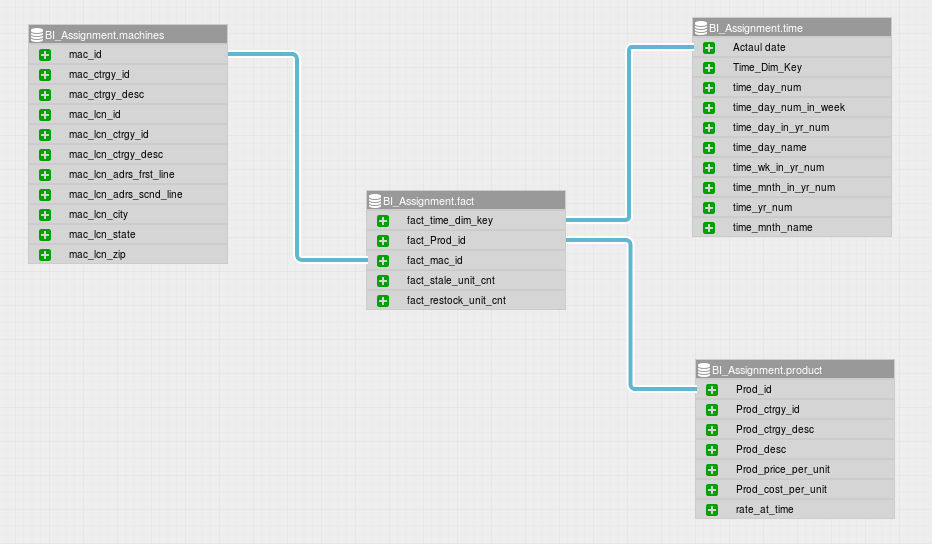
**[Introduction]**

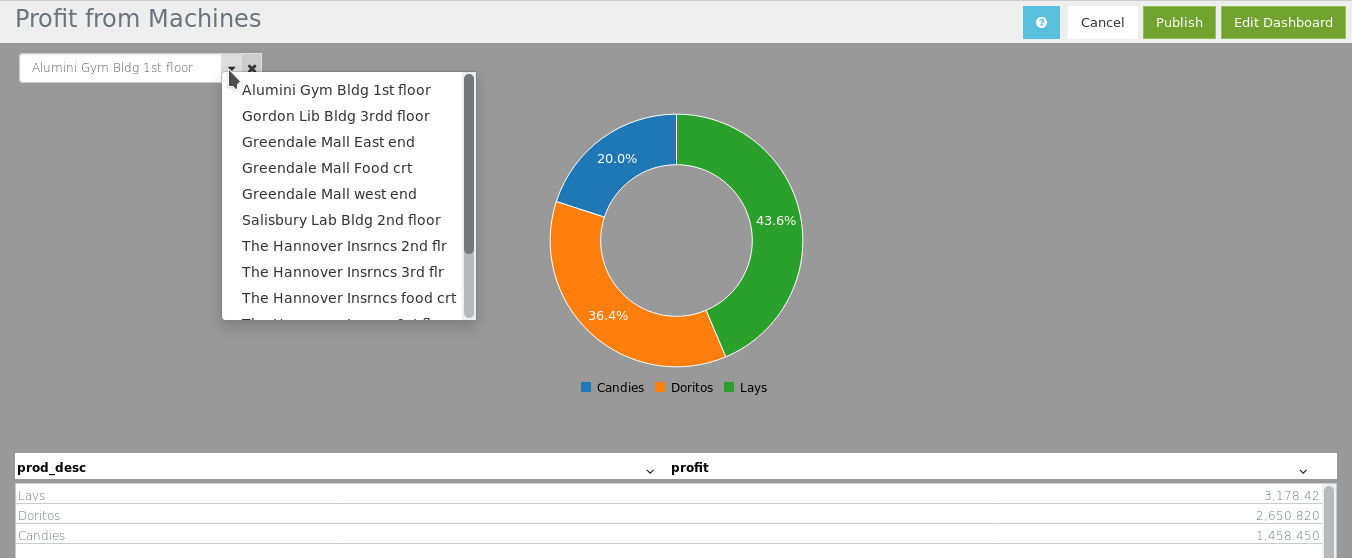
In this paper we are going to perform a profitability analysis for vending machines usage across three different sectors : Schools, Malls, and Companies. We have based our analysis on a sample dataset that should represent a basic overview of the overall distributions of vending machines usages. The main objective of our analysis is to find patterns within the usage recording and draw some conclusions and provide the decision makers with suggestions for better utilization of resources.

**[Dataset]**

The dataset we have used is star schema database with three dimensions: Time, Location and Product. The fact table has some recording for the usage of the vending machine which is the core of this analysis. The usage data stored are the number of items restocked to the vending machine and the number of items that are expired and need to be staled.

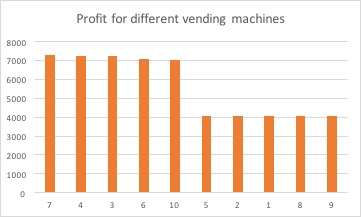


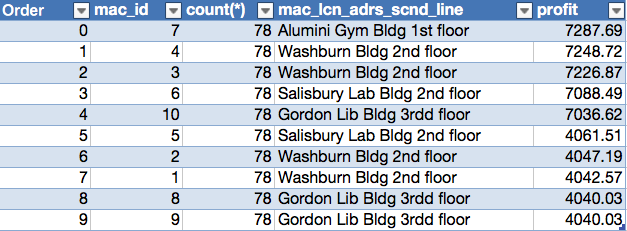
**[Methods]**

In our analysis we have used different methods to analyse the data and gather insights. Essentially, we have used the built in tools within Microsoft Excel software to analyse the data. Also, we have used programming scripting with Python and Pandas Library to perform data analysis. We have used a package called pandasql to incorporate classical SQL query language to query the data frame directly and print out tables and draw charts. To represent the analysis, we created interactive dashboards in a open source BI tool called WideStage. We used a mysql database as the backend for the dashboards. The dashboards would give the flexibility for the manager to drill down to a granular level. For Example, the below dashboard allows the user to select the location and the profit from all the machine at that location is given broken down by the product and their share in the profit for the location. 

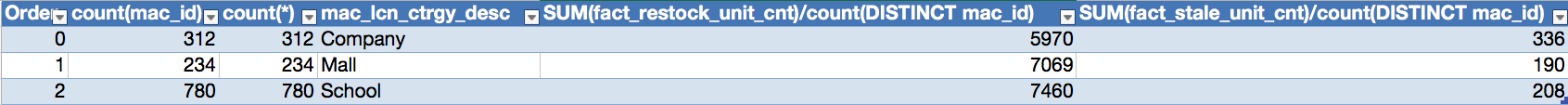
**[Results]**

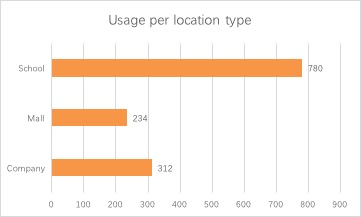
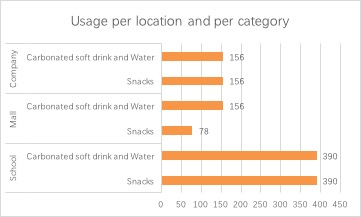
Firstly, we have compared the usage patterns for different vending machines at WPI and measured their profitability turns out that there are three machines with the highest profitability which are one in 1st floor at the Gym and two of them are in the 2nd floor of Washburn Building. On the other hand the machines with least profitability are two in 2nd Floor of the library and one in 2nd floor of Washburn Building. It seems one of the machine at 2nd floor of Washburn Building was making way more money than the other. The reason might be the accessibility of snack vs drinks.



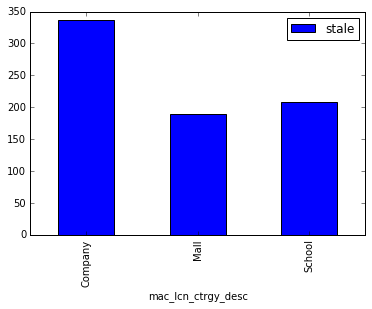


We also analyzed the different usage patterns across all different sectors (WPI school / Greendale Mall and The Hannover Insurance Company) and find some interesting patterns. First when we want to analyse the usage records for these different sectors we had to correct the averaging numbers based on how many machines are in different sector because the number of machines are different. So the total sum values of restock and stale are divided by the number of machines in each sector.



From a side by side comparison, it it clear that school’s usage exceed others’ usage.  
   
What’s more, the chart of usage by location and machine category shows that the usage is equivalent for snacks and drinks machines at school and company. However, the usage at mall shows that people prefer to buy drinks.  


When we looked at the stale patterns across different sectors we found that the stale frequencies are about the same in WPI school and greendale mall. However, the stale amounts are way bigger in Hannover Insurance company.



Thirdly, we analyze the machines differ by the sum number of stales.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| mac\_id | mac\_lcn\_adrs\_scnd\_line | Stales | mac\_id | mac\_lcn\_adrs\_scnd\_line | Stales |
| 2 | Washburn Bldg 2nd floor | 0 | **1** | Washburn Bldg 2nd floor | 11 |
| 5 | Salisbury Lab Bldg 2nd floor | 0 | **7** | Alumini Gym Bldg 1st floor | 241 |
| 8 | Gordon Lib Bldg 3rdd floor | 0 | **4** | Washburn Bldg 2nd floor | 292 |
| 9 | Gordon Lib Bldg 3rdd floor | 0 | **3** | Washburn Bldg 2nd floor | 336 |
| 11 | The Hannover Insrncs1st flr | 0 | **17** | Greendale Mall Food crt | 572 |
| 13 | The Hannover Insrncs3st flr | 0 | **6** | Salisbury Lab Bldg 2nd floor | 603 |
| 15 | Greendale Mall west end | 0 | **10** | Gordon Lib Bldg 3rdd floor | 603 |
| 16 | Greendale Mall East end | 0 | **14** | The Hannover Insrncs food crt | 630 |
|  |  |  | **12** | The Hannover Insrncs2st flr | 717 |

Half of the machines just hold drinks, so their stales are 0. Another half of the machines hold snacks and their stales are also different. Machine 12 and 14 in the company Hannover insurance have the highest stales. Machine 10, 6 in WPI and machine 17 in Greendale Mall also have relatively high stale.

**[Conclusion]**

Based the analysis, we suggest a few changes to improve the profitability.

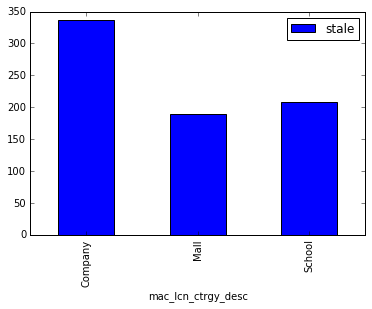
I have three suggestions as follows to improve your profit:

1. For the machines in WPI, Your company should stock more products in the machine 7 (location: Alumini Gym Bldg 1st floor), machine 3 (Washburn Bldg 2nd floor) and machine 4 (Washburn Bldg 2nd floor). These machines are top 3 most profitable. As there is just one machine in the Alumini Gym Bldg 1st floor, the company can load one more machine here. What’s more, you can withdraw machine 1 or machine 2 in Washburn Bldg 2nd floor. The reason is these two machines are the least profitable machines and their sum profit is roughly equal to machine 3 in the same place.
2. You should load more machines and stock more products at schools than companies or malls. School stock many more products than company, mall, and its stale rate is reasonable. Secondly, as the stale rate of Company is much higher than mall and school, you should put less snacks in company than now. Thirdly, the restock number of Carbonated soft drink and Water at malls are much higher than that of snacks. For other malls, you can stock more Carbonated soft drink and Water than snacks in vending machines.
3. You need to stock less snacks to machine 12 and 14 in the company Hannover insurance, machine 10 and 6 in WPI, and machine 17 in Greendale Mall. The reason is that history data shows that these machines have the highest stales.

**Appendix A**  SQL Queries

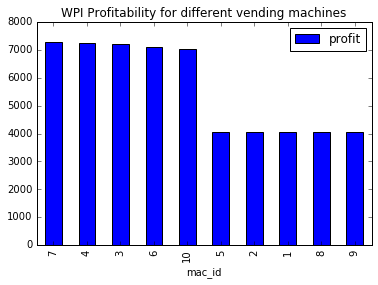
query = "select mac\_lcn\_ctrgy\_desc,SUM(fact\_restock\_unit\_cnt)/count(DISTINCT mac\_id) as stock " \  
 ",SUM(fact\_stale\_unit\_cnt)/count(DISTINCT mac\_id) as stale from df group by mac\_lcn\_ctrgy\_desc "   
df2 = pysql(query)  
df2.plot(kind='bar',x = "mac\_lcn\_ctrgy\_desc", y=["stale"])

<matplotlib.axes.\_subplots.AxesSubplot at 0x117a08bd0>



str2 = "select mac\_id,count(\*),mac\_lcn\_adrs\_scnd\_line, " \  
 "SUM( (fact\_restock\_unit\_cnt\*Prod\_price\_per\_unit) - " \  
 "((fact\_restock\_unit\_cnt+fact\_stale\_unit\_cnt)\*Prod\_cost\_per\_unit )) as profit " \  
 "from school group by mac\_lcn\_adrs\_scnd\_line,mac\_id order by profit desc "  
profit\_df = pysql(str2);  
profit\_df.plot(kind='bar',x = "mac\_id", y=["profit"],title = "WPI Profitability for different vending machines")

<matplotlib.axes.\_subplots.AxesSubplot at 0x118367610>



str2 = "select mac\_lcn\_ctrgy\_desc,mac\_lcn\_ctrgy\_id ,mac\_id,count(\*),mac\_lcn\_adrs\_scnd\_line, " \  
 "SUM( (fact\_restock\_unit\_cnt\*Prod\_price\_per\_unit) - " \  
 "((fact\_restock\_unit\_cnt+fact\_stale\_unit\_cnt)\*Prod\_cost\_per\_unit )) as profit " \  
 "from df group by mac\_lcn\_adrs\_scnd\_line,mac\_id,mac\_lcn\_ctrgy\_desc,mac\_lcn\_ctrgy\_id order by profit desc "  
profit\_df = pysql(str2);  
  
   
profit\_df.plot(kind='bar',x = ["mac\_id","mac\_lcn\_ctrgy\_desc"], y=["profit"],title = "WPI Profitability for different vending machines")

