

Report of DW Assignment

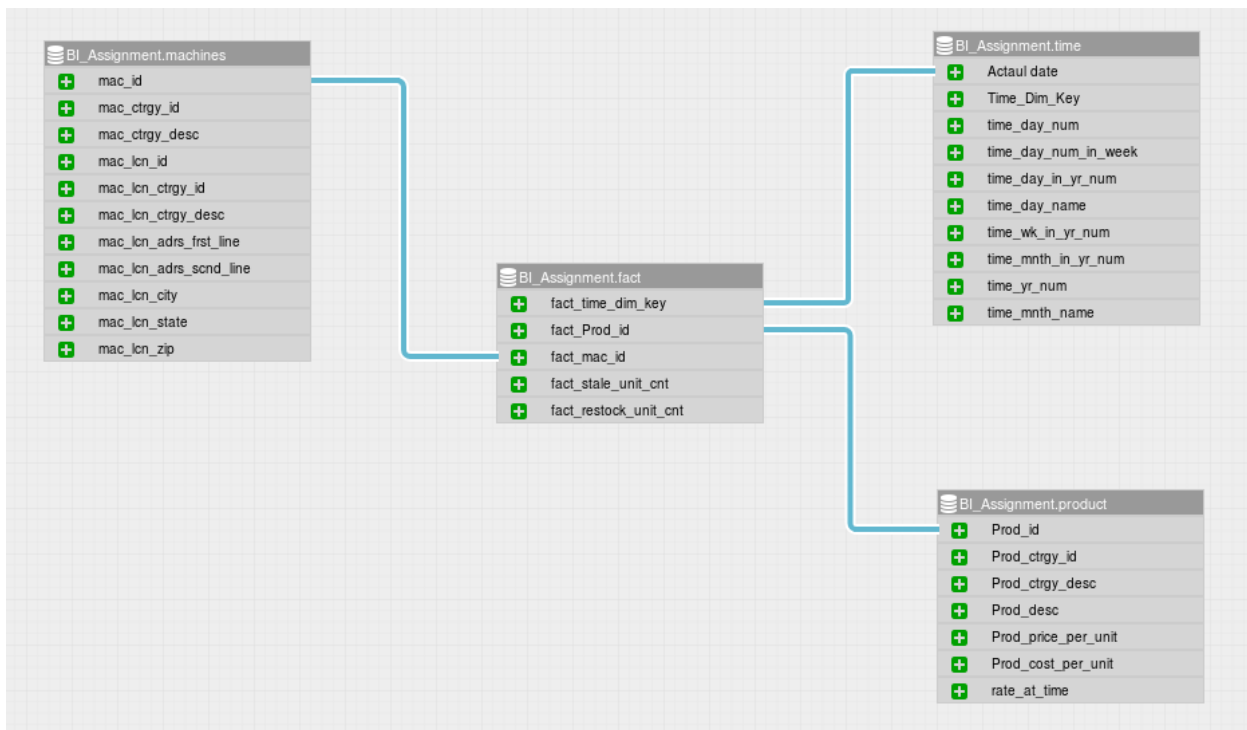
Kavin Chandrasekaran, Abdulaziz Saleh, Fangling Zhang and Sishi Hao

[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.

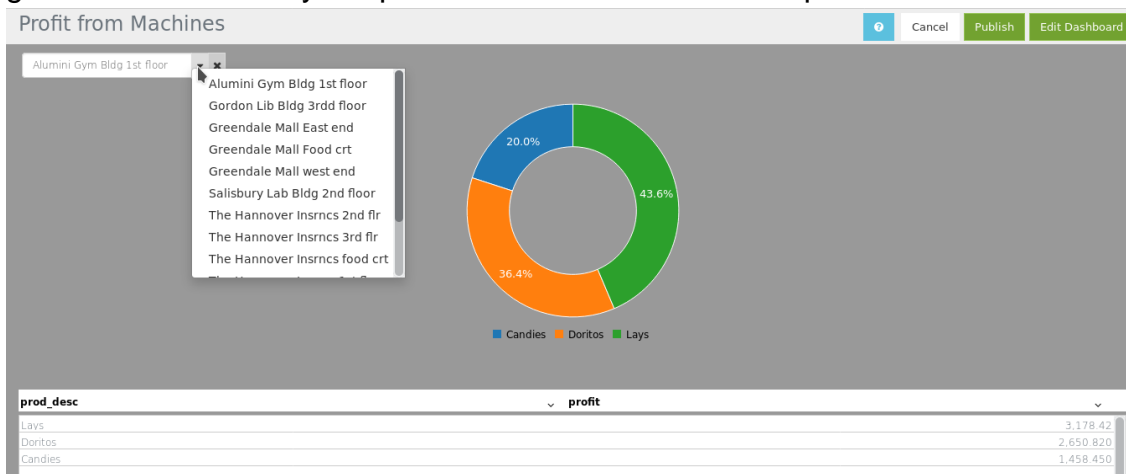
[Dataset]

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 stale.



[Methods]

In our analysis we have used different methods to analyze 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]

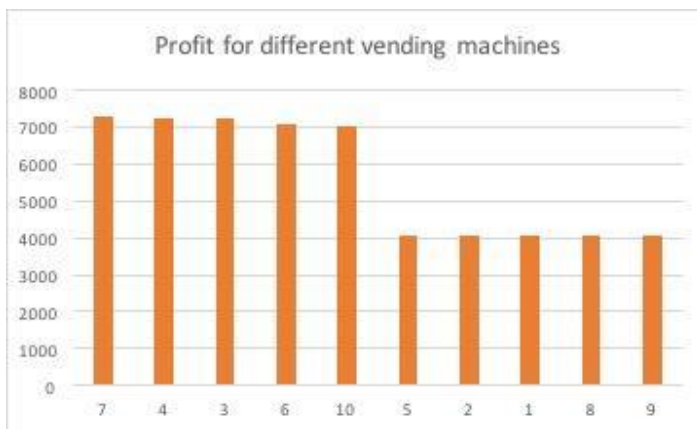
Q1. Profit analysis for vending machines at WPI

Firstly, we have analyzed profitability of the 10 different vending machines at WPI. The process of profit calculation is as follows: Firstly, for each machine, each product and each restocking time, we deducted the number of stale from the number of new items that were put into the machine last time, so we got the quantity sold in a small period. Secondly, we multiplied the quantity with $(\text{Prod_price_per_unit} - \text{Prod_cost_per_unit})$, thus we get the profit in a small period for each product and each machine. Lastly, we summed up the profit together for each machine.

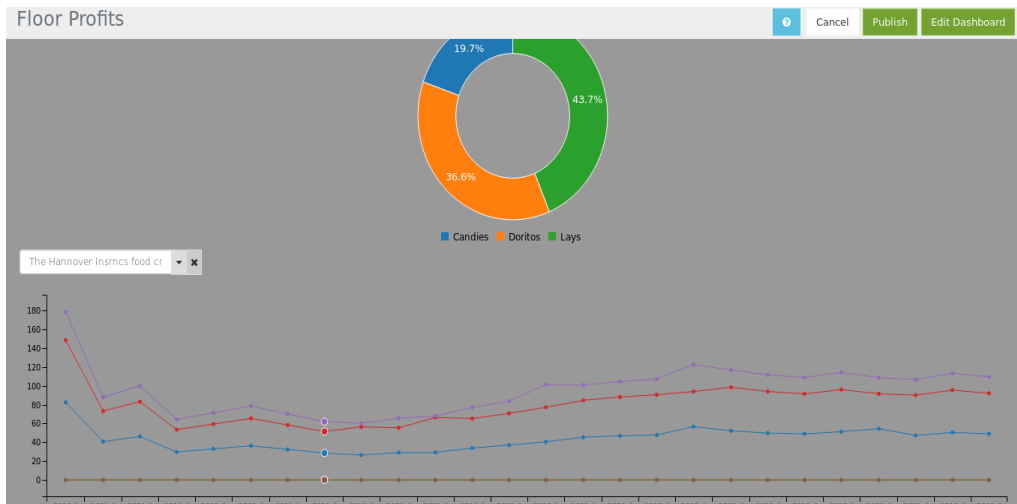
The results show that top 3 machines with the highest profitability are one in 1st floor at the Gym and two of them are in the 2nd floor of Washburn Building. On the other hand the machines with least profitability are two in 2nd Floor of the library and one in 2nd floor of Washburn Building. It seems 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.

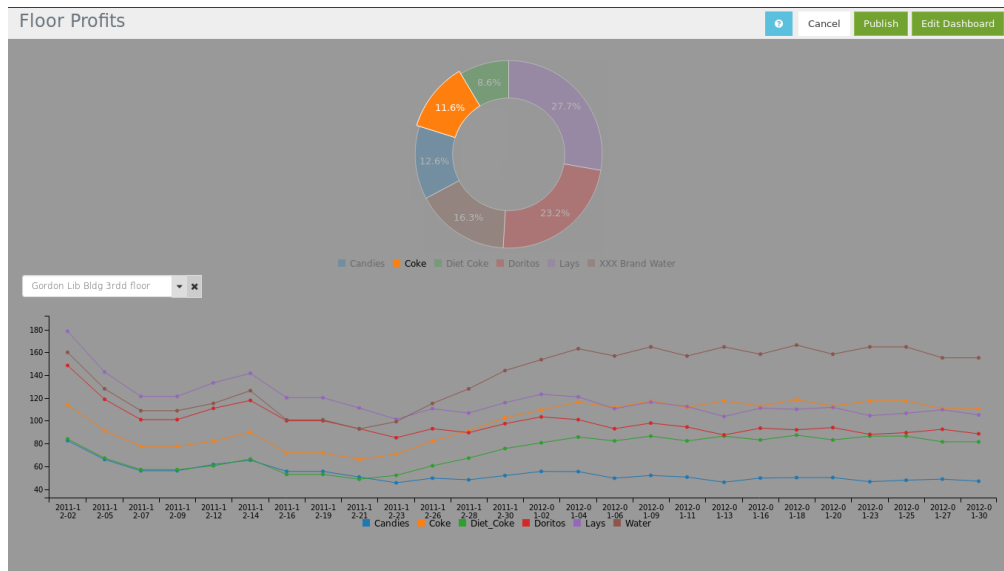
Order	mac_id	count(*)	mac_lcn_adrs_scnd_line	profit
0		7	78 Alumini Gym Bldg 1st floor	7287.69
1		4	78 Washburn Bldg 2nd floor	7248.72
2		3	78 Washburn Bldg 2nd floor	7226.87
3		6	78 Salisbury Lab Bldg 2nd floor	7088.49
4		10	78 Gordon Lib Bldg 3rdd floor	7036.62
5		5	78 Salisbury Lab Bldg 2nd floor	4061.51
6		2	78 Washburn Bldg 2nd floor	4047.19
7		1	78 Washburn Bldg 2nd floor	4042.57
8		8	78 Gordon Lib Bldg 3rdd floor	4040.03
9		9	78 Gordon Lib Bldg 3rdd floor	4040.03



We also analyzed profit trend for the library building in WPI with school events.



The above dashboard shows varying profit trend for the library building in WPI and this could be associated with school events but the below dashboard shows and office and trend is fairly consistent. By considering special events, the restocking frequency can be adjusted to minimize operating costs.



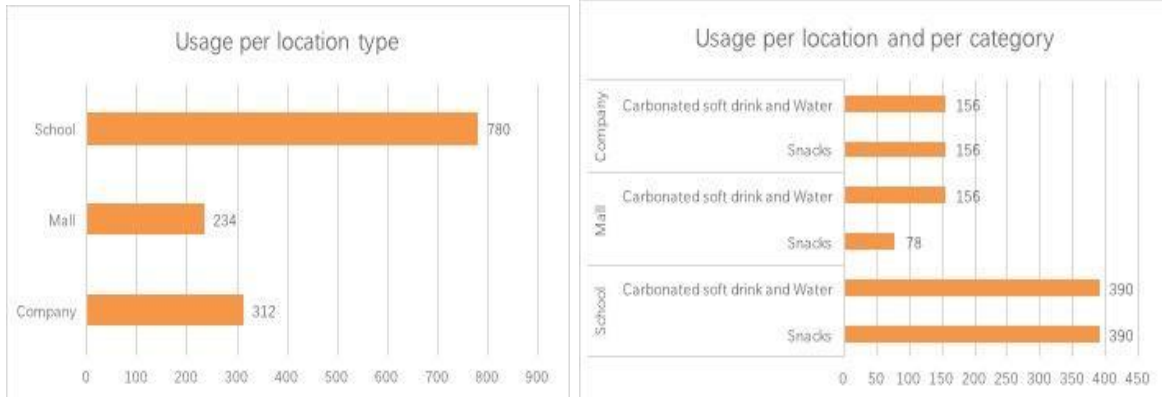
Q2. The usage (represented by stocking and stales) differ for machines at malls vs. companies vs. schools (WPI).

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.

<u>mac lcn ctrgv desc</u> <u>/Prod ctrgv desc</u>	Sum of <u>fact stale unit cnt</u>	Sum of <u>fact restock unit cnt</u>	Stale_ rate
Company	1347	23883	5.64%
Carbonated soft drink and Water	0	10596	
Snacks	1347	13287	
Mall	572	21207	2.70%
Carbonated soft drink and Water	0	13302	
Snacks	572	7905	
School	2086	74604	2.80%
Carbonated soft drink and Water	11	33915	
Snacks	2075	40689	

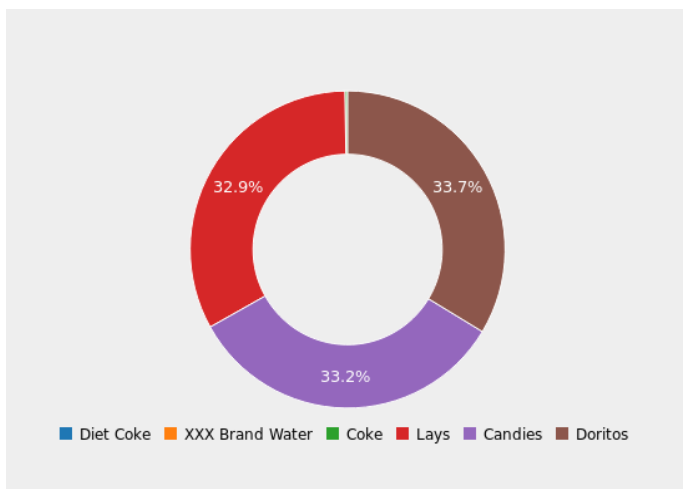
From a side by side comparison, it is clear that school's usage exceed others' usage.

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 drink.



When we looked at the stale patterns across different sectors we found that the stale rate (stale_unit_cnt/restock_unit_cnt) are about the same in school and greendale mall. However, the stale rate is much bigger in company.

When we looked at the stale patterns across different products, we found that doritos has the highest number of stale units left.



Q3. Machines differ by the number of stale.

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

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]

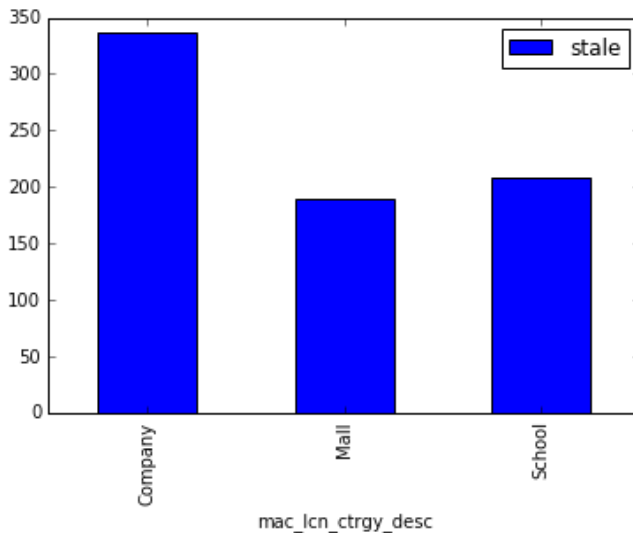
Based the analysis, we suggest a few changes to improve the profitability.

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. We would also suggest to consider the seasonality in restocking. For example, when the school closes during the winter break the restocking could be reduced contrary to the regular school days. But this wouldn't apply to public places like the mall and offices.
2. If you plan to load machine in other cities, you should load more machines and stock more products at schools than companies or malls. School restock 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 (Hannover) 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. Fourthly, Based on looking at the number of stales across all products, we see that doritos has the highest number of stale units left. We would suggest decreasing the number of doritos restocked in the machine, so that the number of stale units can be reduced and the supply can be increased based on demand.
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 stale.

Appendix A SQL Queries

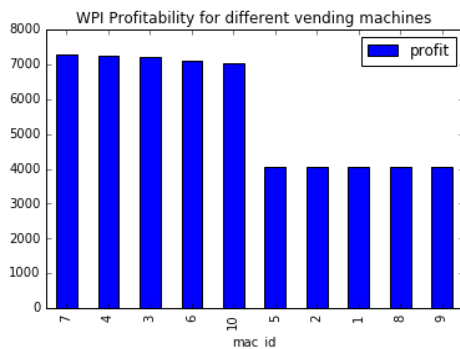
```
query = "select
mac_lcn_ctrgy_desc,SUM(fact_restock_unit_cnt)/count(DISTINCT mac_id) as
stock " \
",SUM(fact_stale_unit_cnt)/count(DISTINCT mac_id) as stale  from df
group by mac_lcn_ctrgy_desc "
df2 = pysql(query)
df2.plot(kind='bar',x = "mac_lcn_ctrgy_desc", y=["stale"])
```

<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>



```

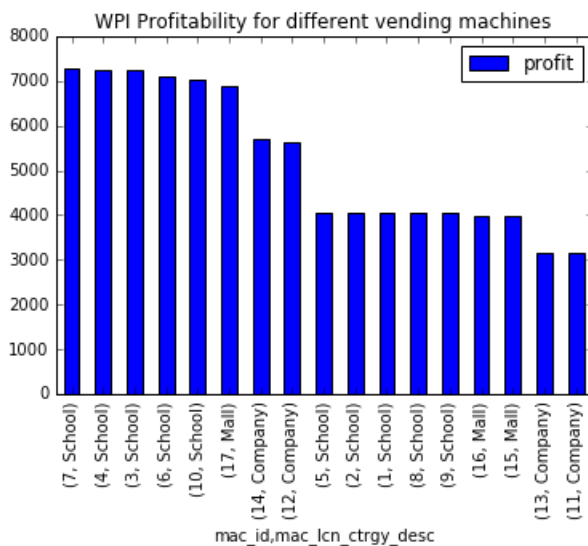
str2 = "select mac_lcn_ctrgy_desc,mac_lcn_ctrgy_id
        ,mac_id,count(*),mac_lcn_adrs_scnd_line, " \
        "SUM( (fact_restock_unit_cnt*Prod_price_per_unit) - " \
        "((fact_restock_unit_cnt+fact_stale_unit_cnt)*Prod_cost_per_unit ))
as profit " \
        "from df group by
mac_lcn_adrs_scnd_line,mac_id,mac_lcn_ctrgy_desc,mac_lcn_ctrgy_id
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")

```



```

select Floor, Category, DATE_FORMAT(`Date`, '%Y-%m-%d') as `Date`,
SUM(Candies) as Candies,
SUM(Coke) as Coke,
SUM(Diet_Coke) as Diet_Coke,
SUM(Doritos) as Doritos,
SUM(Lays) as Lays,
SUM(Water) as Water
from
(
select Machine, Floor, Category, `Date`,
if(Product="Candies", Percentage , 0 ) as Candies,
if(Product="Coke", Percentage , 0 ) as Coke,
if(Product="Diet Coke", Percentage , 0 ) as Diet_Coke,

```



```

if(Product="Doritos", Percentage , 0 ) as Doritos,
if(Product="Lays", Percentage , 0 ) as Lays,
if(Product="XXX Brand Water", Percentage , 0 ) as Water from
(select prod_profit.mac_id as Machine,Floor, prod_profit.mac_lcn_ctrgy_desc
as Category, prod_profit.`Date` as `Date`, prod_desc as Product,
prod_profit as Percentage from
(select mac_id, mac_lcn_ctrgy_desc, mac_lcn_adrs_scnd_line as Floor,
STR_TO_DATE(left(`Actaul Date`,length(`Actaul Date`)-4), '%m/%d/%Y') as
`Date`, prod_desc,
round(SUM( (fact_restock_unit_cnt*Prod_price_per_unit) -
((fact_restock_unit_cnt+fact_stale_unit_cnt)*Prod_cost_per_unit )),2) as
prod_profit
from vending_machines group by mac_id, Floor, Prod_desc, mac_id ,
mac_lcn_ctrgy_desc, `Actaul Date`) prod_profit
) floor_group
) cat_group
group by Floor, Category, `Date` order by `Date` asc

select mac_profit.mac_id, prod_profit.prod_desc,
round((total/(sum(prod_profit.profit)))*100,2) as percentage from
(select mac_id,
round(SUM( (fact_restock_unit_cnt*Prod_price_per_unit) -
((fact_restock_unit_cnt+fact_stale_unit_cnt)*Prod_cost_per_unit )),2) as
total
from vending_machines group by mac_id) mac_profit left join
(select mac_id, prod_desc,
round(SUM( (fact_restock_unit_cnt*Prod_price_per_unit) -
((fact_restock_unit_cnt+fact_stale_unit_cnt)*Prod_cost_per_unit )),2) as
profit
from vending_machines group by mac_id, Prod_desc) prod_profit
on mac_profit.mac_id=prod_profit.mac_id
group by mac_id, prod_profit.prod_desc;

```