

**ASSIGNMENT FRONT SHEET**

**Course Name: ALY6050 20906 Intro to Enterprise Analytics**

**Professor Name:**  **Dr. Christopher Johnson**

**Student Name: Dong Quoc Tuong (Lukas), Meghana Musunoori, Anita Parmar**

**Student Class: Fall 2019 CPS Term: B. 2020**

|  |
| --- |
| **Signature Project- Sales Performance Analysis**  **Completion Date:t Due Time:12:00am** |

**Statement of Authorship**

*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s regulations to copy the work of another without clear acknowledgement and that attempting to do so renders me liable to disciplinary procedures. To this effect, I have uploaded my work onto Turnitin and have ensured that I have made any relevant corrections to my work prior to submission.*

**Tick here** to confirm that your paper version is identical to the version submitted through Turnitin

**Questions Asked**

Below are some of the questions we asked about the data when we started to explore it:

* Are there any anomalies in the data that we need to remove before doing analysis on it?
* Is there a larger concentration of members in one part of the country over another?
* Does time of year impact the fluctuations of sales data, holiday, closures?
* Who is our audience that we are presenting to and how will we change the view based on the audience and the story we want to tell them?

**Data Sources**

* D100-Sales- FY19P1to7 from Costco

**Variables**

The following variables were taken into account while analyzing this dataset:

* US Holidays

**Assumptions**

While analyzing the data, we noticed an anomaly with the data in Register = 999. The data for this criterion did not match the patterns of the rest of the dataset. Based on this, we concluded it was best to exclude where Register = 999 in order to get more accurate patterns from the rest of the dataset.

**Methods**

We used the following Tableau graphs to analyze this data:

* Geographical maps
* bar charts
* line graphs

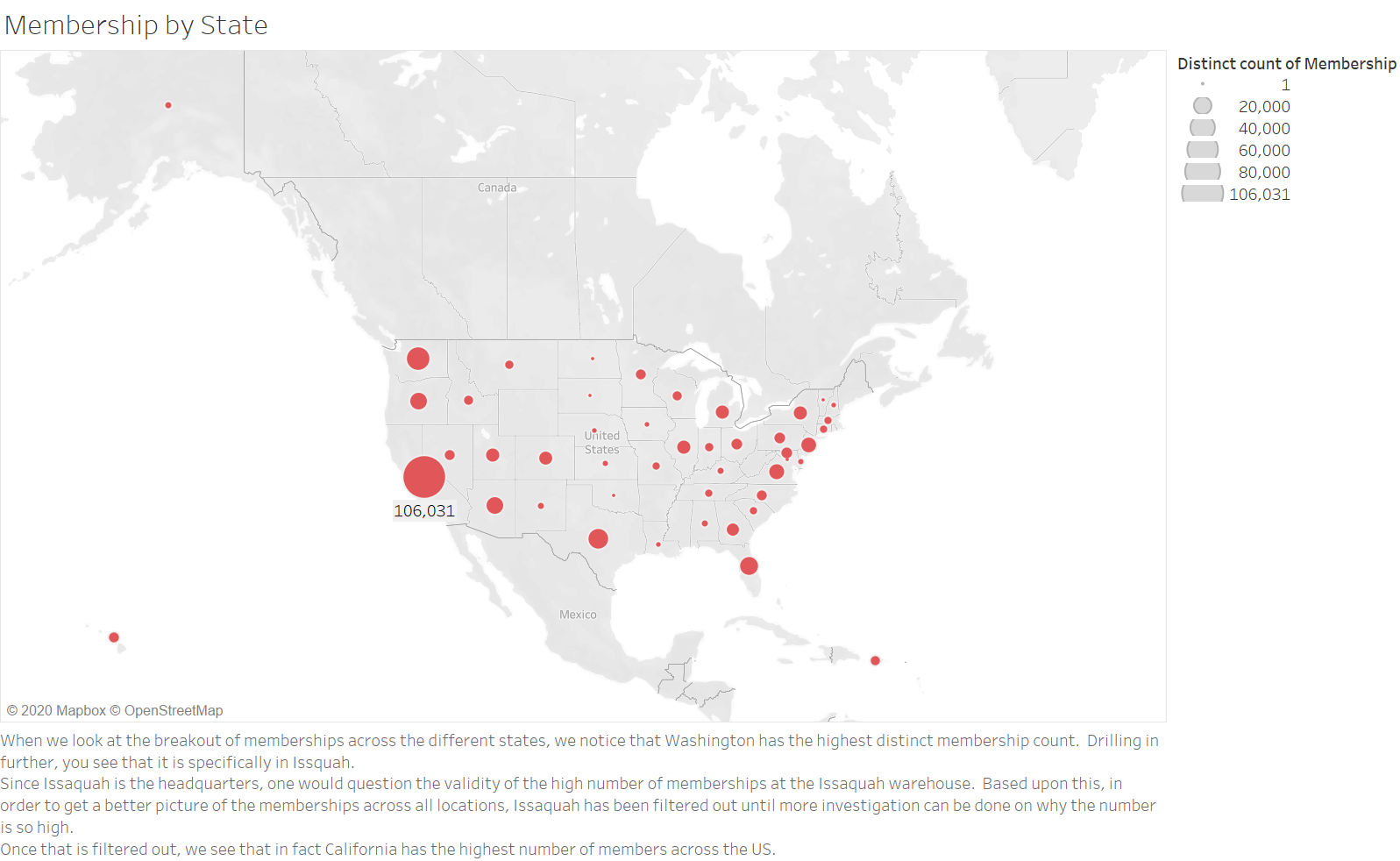
In addition, we found that creating a hierarchy from State 🡪 City 🡪 Warehouse 🡪 Register was a good way to break down the data into more detail.

**PART – A:**

**Findings**

1. **Memberships at Costco differ by the state you are in.**

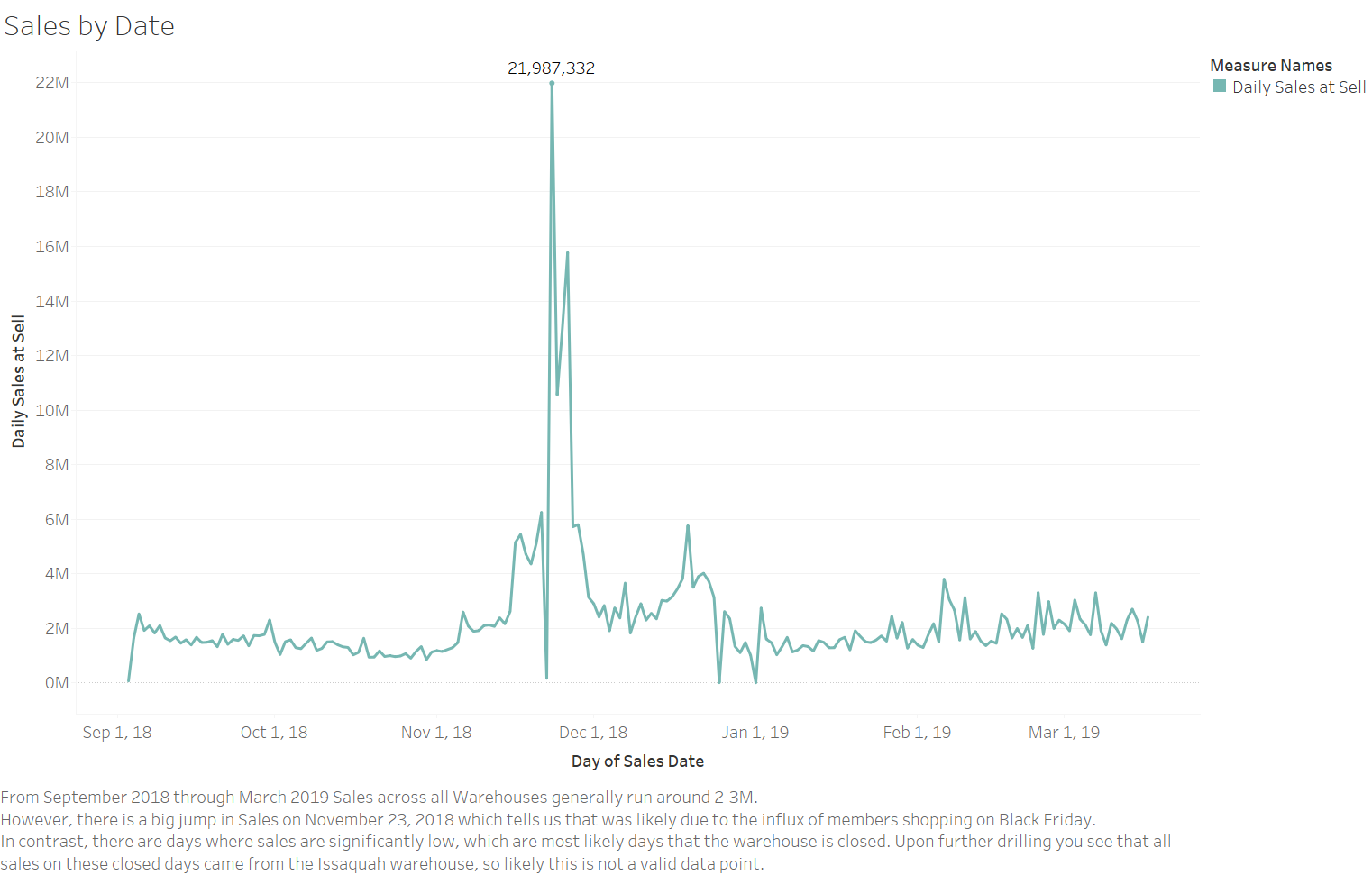
Per the graph below, we found that once Issaquah was excluded as a city, California had the highest number of memberships and South Dakota and Oklahoma had the lowest. Another interesting finding was that Arkansas did not have any memberships. Is this due to it being Walmart’s headquarters, so there is no Costco in that state?



1. **Sales data is fairly consistent across months other than for Holidays**

Generally speaking, it appears that Sales data stays at around 2M consistently. The Black Friday weekend and lead up to, shows the peak of sales, with close to Christmas time as the second busiest time.

Days where there are close to no sales are days were the warehouses were closed.



1. **Sales and refunds by each month**

A line graph representing the amount of sales for each month is created by using sales date in terms of months as dimension and the sum of Daily units at sell as the measure.

A close up of a map

Description automatically generated

From the above graph we can observe that Costco has highest sales in the month of November i.e. $65,407,927 and lowest sales in the month of March i.e. $18,919,365. The drastic increase of sales in the November might be due to Black Friday and Christmas sales.

A picture containing screenshot

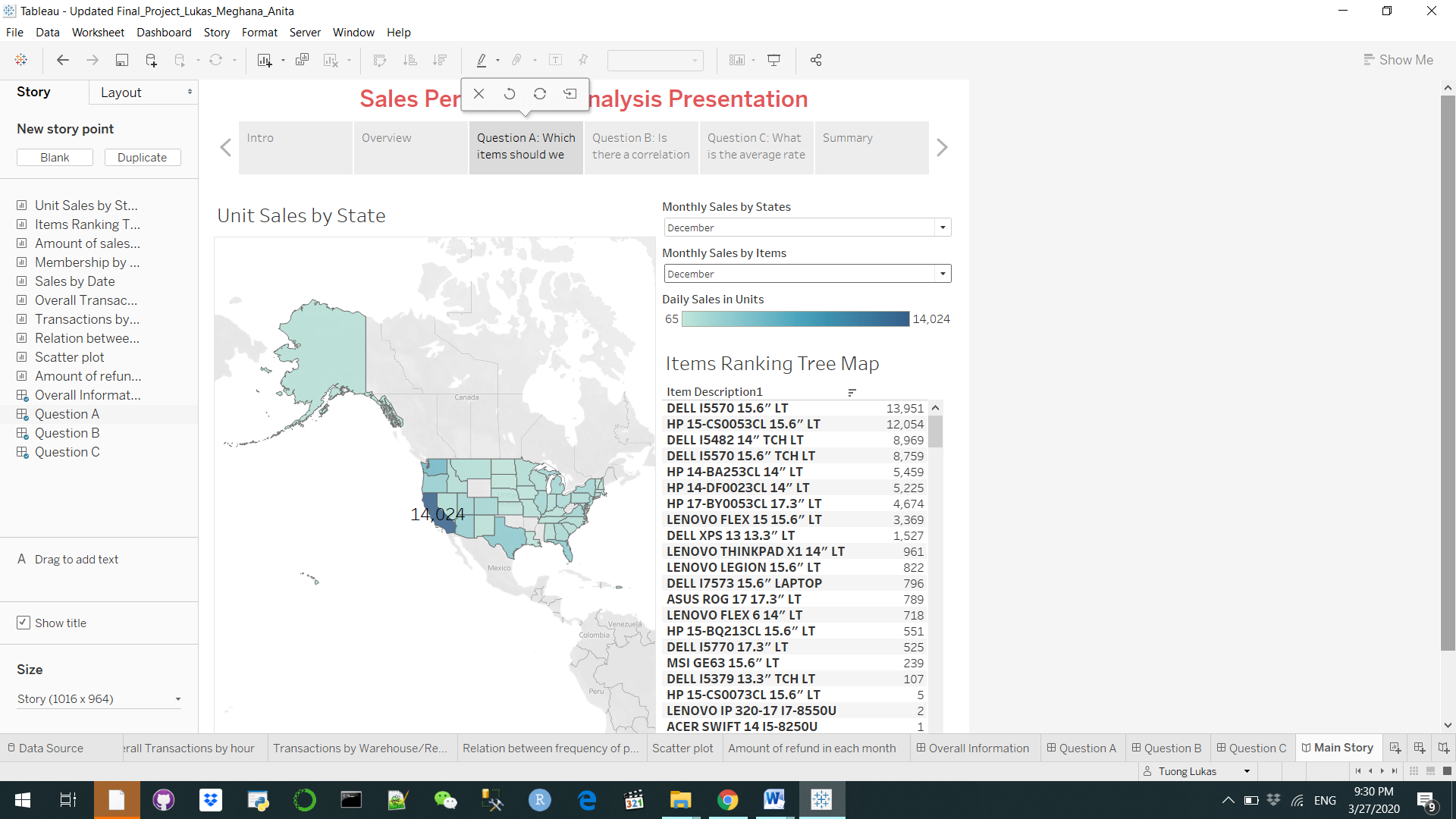
Description automatically generated

* The amount of refund in each month graph is created by using the Daily sales in units as a dimension where all the units that are negative are grouped together as Refund items and all the units that are positive are grouped together as Purchased items.
* The sales date in terms of months and the refund flag are also used as dimensions and the sum of Daily sales at sell is used as the measure.
* The daily sales in units is filtered to select only refund items and refund flag is filtered to select only “Y” i.e. the items for which the refund key is pressed
* From the resultant graph we can see that December has highest amount of refunds i.e. $19,815,687 and March has the lowest amount of refunds i.e. $6,385,207.

**PART – B:**

1. **Stock suggestions by region and time**

* In order to find the amount of stocking that we need to do per item per state in different months, we assign 2 maps (geographic and tree map) and then link them together setting the interactive mode
* The Daily unit sales are demonstrated by the heat chart and we must make sure that the 2 monthly filters have to match each other in order to get the right result
* From a respective point of view, California has the highest number of units sold in every month (from 8000 to approximately 15000 units per day). This followed by other states in the West coast. Midwestern and East coast states have the lowest number of daily sales in unit. Additionally, from the Item Ranking tree map, we see that Dell and HP computers remain best sellers for the company. Particularly Dell I5570 15.6 and HP 15-CS0052CL contributed to more than half of the average unit sales each month.



1. **Correlation between the frequency of purchase of any item and the region or the city of purchase**

* In order to find if there exist any correlation between the frequency of purchase of any item and the region, we first calculated the frequency of sales per item and region

}

* From the above two measures we can now plot a scatter plot which shows any correlation that exists between

frequency of purchase of any item and region.

* We have excluded the area EC from the scatter plot by filtering the Register Number 999 based on the assumption that it represents E-Commerce and doesn’t particularly pertain to any region.

A picture containing white, black, bus

Description automatically generated

* From the above scatter plot, we can observe that there exists no correlation between the frequency of purchase of any item and region. One interesting relationship we found from the scatter plot is that the regions with highest sales also have the highest frequency of purchase

A screenshot of a cell phone

Description automatically generated

* Now from both the scatter plot and bar graph we can observe that the region NW which has the highest daily sales also has the highest frequency of purchase per item
* At the same time the region BD which has the lowest daily sales also has the lowest frequency of purchase per item

1. **Average Transaction Time by Warehouse**

Overall the average transaction time by warehouse generally runs at about 1500 seconds with the highest time being at Rego Park and the lowest time being at Commerce. This data can then be broken out by register by hour for further details

