Datathon Submission Template

Retina AI (https://retina.ai)

24 September, 2020

Retina AI R Datathon - Submission Template

Team Name: Zoom University
Team Members: Kyle Fang, Jiahao Huang, William Foote
Team Members Email: kylefang19@g.ucla.edu (mailto:kylefang19@g.ucla.edu), jiahao303@g.ucla.edu (mailto:jiahao303@g.ucla.edu), williamfoote@g.ucla.edu (mailto:williamfoote@g.ucla.edu)
Designated Slides: Slides 12-13

Directions: Write your text and codes for Task 1, 2 and 3 in the provided space below.

Task 1 insights to the visualization can be written as plain text in the field provided below. You can also add new bullet points to each slide's insights section.

For Task 2 and 3, provide your codes in the code chunks below. If your team needs additional code chunks to run your code, you can add new code chunks.

Task 1 (Required)

Slide Number: 12

Key Insights min 2 / max 4:
1. The tailedness of a distribution quanitifed by kurtosis. A heavy tail in the raw number of customers indicates a very highly peaked distribution or many outliers. The implications of this are dependent on the skewedness.

A heavy tail with left-skewed data translates to a lot of high-paying customers, which is very good for the company. However, a heavy tail matched with right-skewed data translates to reliance on a few high-paying customers.

On the other hand a low tail indicates a flatter distribution and less outliers. Generally there is less risk as the customer based is diversified and not heavily relying on a few high-paying customers.

2. The skewedness of the plot measured by skewedness. A left skew means there are a lot of high-paying customers, which is great for the business. No skew means the company is reliant on the average customer, and the average customer is robust. Right skew means there are only a few high-paying customers, which is not good because the company is reliant on these few high-paying customers.

Slide Number: 13

Key Insights min 2 / max 4:
1. Whether the mean is greater than the median.

If the mean is greater than the median, the LTR distribution is right-skewed. The business is reliant on a small amount of high-spending customers.

If the median is greater than the mean, the LTR distribution is left-skewed. The business has a big number of high-spending customers.

2. Whether the relationship between the median and mean change over time. This has different implications depending on whether or not the mean is greater than median.
3. The existence of “pesky trends”. The periods of time (6 months) where the trend of the difference between the median and mean of the LTR is unusual compared to the overall trend. This has different implications depending on whether or not the mean is greater than the median. The pesky trends are helpful in determining if the trends in the current times are consistent with the overall trend.
4. The exact times of those pesky trends. Those times should be further investigated as it could indicate times of unusual success or failure for the business, and one might be interested to see what other factors lead to these abnormalities.

Task 2 (Required)

Slide Number: 12

```
## [1] "Since the excess kurtosis is high/above 2 (23.29), there is a lot of variance in
the customer lifetime revenues, reflecting a lack of a diverse customer risk profile. Th
e skewness above 1 (4.36) reflects a right-skewed customer lifetime revenue distributio
n, indicating that the presence of many loyal customers (whales) bring the mean up."
```

Slide Number: 13

```
## [1] "The mean is greater than the median, but the mean is converging towards the medi
an. The distribution of LTR is becoming less right-skewed. The business is becoming less
reliant on whales by building a more robust, average customer -- it is removing outliers
that spend the most. The business's risk profile is decreasing, as it becomes less depen
dent on a few, high-spending customers."
```

```
## [1] "There are 7 points of interest to further investigate. The slope estimate (of th
e difference between mean and median LTR) for at least one subset of the data was outlyi
ngly different than the slope estimates of the other sub-groups. The sub-groups beginnin
g at indices  1, 2, 4, 5, 7, 8, 9  should each be investigated for further insight as to
why the trend might have been different for this time period as opposed to the rest of t
he data. What factors may have led to unusual success or failure over a given stretch?"
```

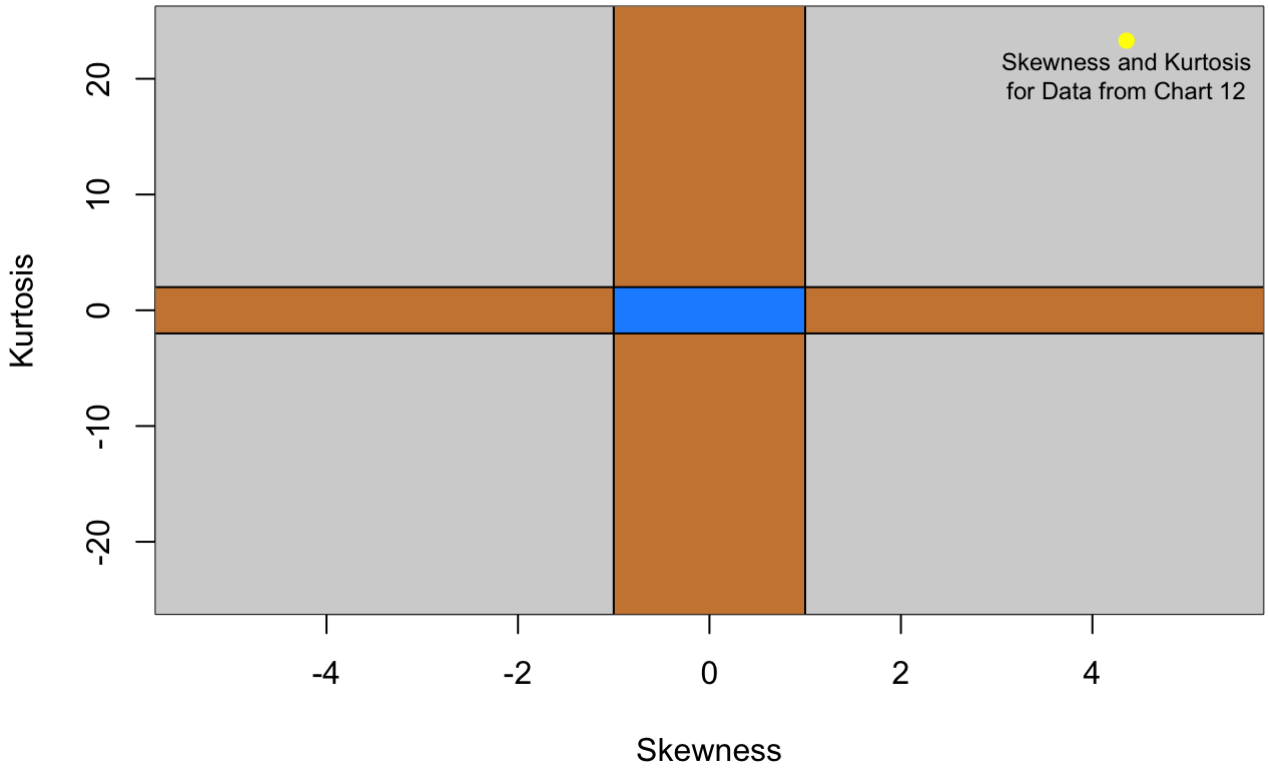
```
## [1] "The recent trend of the LTR matches the overall trend of the data."
```

Task 3 (Bonus/Optional)

Key Insights min 2 / max 4: - Kurtosis and Skewness visualization

How to read this chart: The boxes represent how close or far the point is to being in the ideal kurtosis and skewness levels. This graph visualizes insight 1 and 2 in slide and allows the user to tell how far and in what direction is he away from normal distribution.

Because there is no standard cutoff for kurtosis or skewness in determining whether a distribution is truly classified as normal, this plot can help users get a sense of how far there distribution is from normal distribution.



Key Insights min 2 / max 4: - Boxplot of rolling slope estimates for the difference of the means and medians. *How to read this chart*: - The boxplot shows potential instances where the slope estimate for a given time period subset is outlyingly different than the other slope estimates. It shows visualizes how extreme the outlying are.

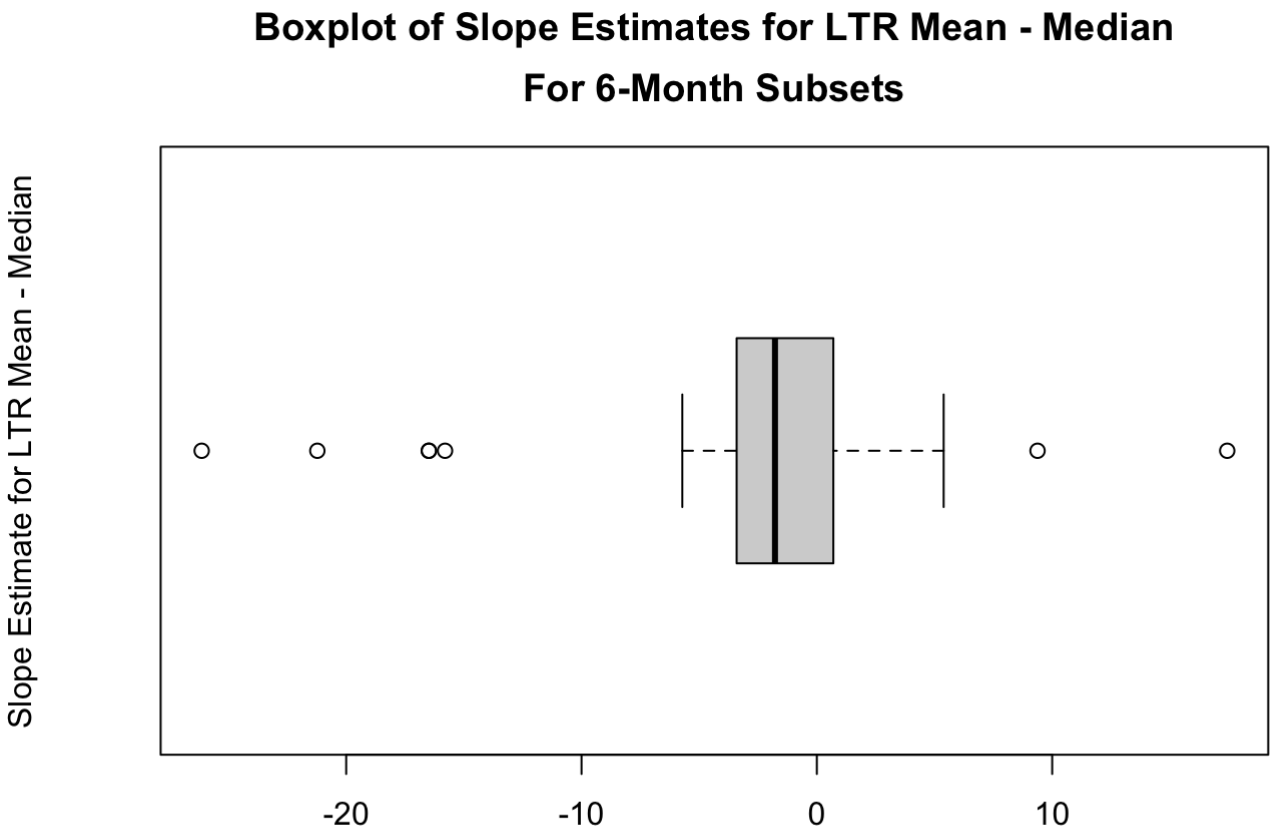
When the mean is greater than the median, the following statements are true.

- The outliers on the right indicate time periods where the businesses are becoming unusually more dependent on a few, high-spending customers.
- The outliers on the left indicate time periods where businesses are unusually successful in becoming less dependent on a few, high-spending customers.

When median is greater than the mean, the opposite is true.

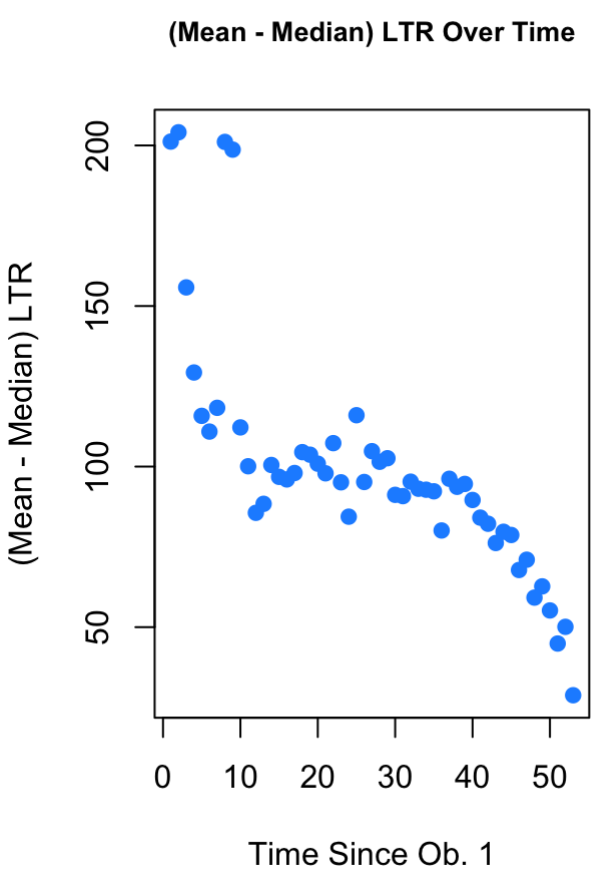
- The outliers on the right indicate time periods where businesses are unusually successful in becoming less dependent on a few, low-spending customers.

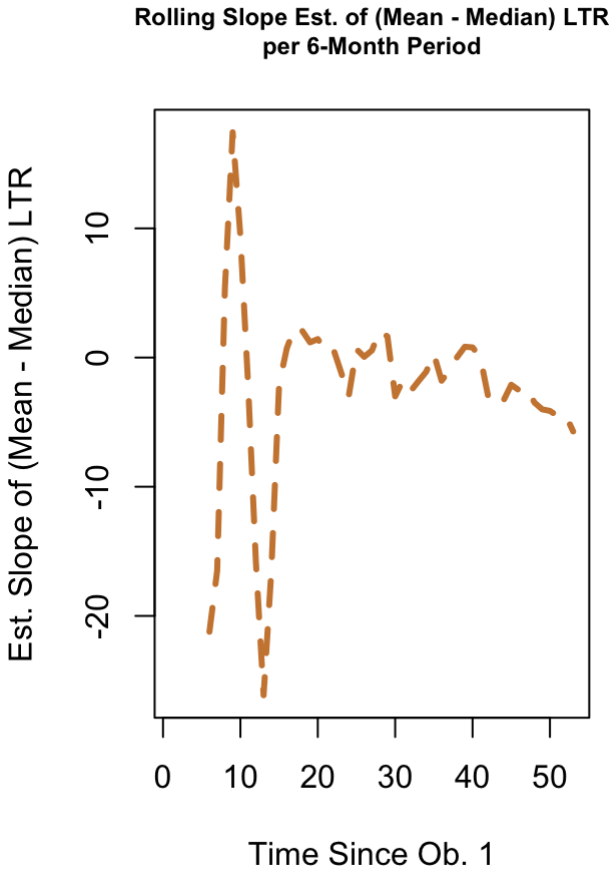
- The outliers on the left indicate time periods where the businesses are becoming unusually more dependent on a few, low-spending customers.



Key Insights min 2 / max 4: - Plots of rolling slope estimates for the difference of the means and medians. *How to read this chart*: These graphs plot the rolling means against time.

If the trend at the end (near current months) is different from the overall trend, there is cause for worry. It should be investigated further as the business's risk profile may have changed recently and could be changing changing currently.


(Mean - Median) LTR Over Time
(Mean - Median) LTR
200
150
100
50
0
10
20
30
40
50
Time Since Ob. 1


Rolling Slope Est. of (Mean - Median) LTR
per 6-Month Period
Est. Slope of (Mean - Median) LTR
10
0
-10
-20
0
10
20
30
40
50
Time Since Ob. 1