Postie Data Challenge Description and Goals:

Feedback

The scenario said it’s July 3rd, 2017, and the new analyst has done his report. The data from the dataset on that day have transactions with timestamps that started at 00:00AM ended at 23:59:38 PM. Was it a typo error?

Question 1:

**1-The analyst wants you to figure out why their 2017-07-03 sales is much than their previous day sales summary**

The company experienced fewer visits compare to the preview days (9981 versus 11573 on July and 11634 on July 1st). Because July 3rd is a federal holiday several reasons can explain the lower number of visits (people are travelling, in vacation, have already made the necessary shopping, …)

**2-The new analyst would also like a report detailing other key metrics for the system. Other that the average sales per day, are there any other metrics that they should be using? Is an average sales value the right metric to use? Explain why /why not?**

I would not ask the question this way. I would rather ask if the average daily sales is enough. Any metrics, if it can help improve performance is good, but no metrics will be good enough.

Some metrics suggestions:

It will be good also the total number of people visiting the website and being able to calculate the conversion (number of checkouts by number of website visits)

Time spent on the website and compare app version1 versus app version 2

What the average revenue per visitor (ARPV) is also a good metrics to have

Can you the feedback when the customer is on the site (their feedback, what they think we can do to implore their experience, interactions with the website, what they like or dislike)

I calculated the ‘average revenue per visitor’ and found that July 3rd performance is far better the 2 previous days

**3-What information can you extract from the urls? Can you infer all product prices? Is there any other information that you believe would be useful to understand what what is going on?**

The url lists the product that have been checked out and it is possible to calculate the price of the product. By computing the value count of the url, one should able to determine the most profitable/wanted product. I have listed the 10 most ‘ checked out urls’. One can also add the timestamp for a better understanding of these products

**4-Are there any interesting purchasing combinations, events, or metrics that are worth reporting and displaying? What information should the analyst know about the system that you’re uncovered?**

The average revenue per visitor

The most popular product

The ‘ effect of app version

**5-Can you predict the total sales number (in dollars) for 2017-07-04? How certain is the predicted number? State what the prediction is doing and what general steps you did to get the number you report and what assumptions you have made**.

The dataset is too small for a good prediction. First I noticed that some timestamp were duplicated, so I have to groupy by timestamp. Checking for frequency, I did not find one. I set the frequency to daily which gives a dataset of length 3. 2 days were used for training and the last for prediction. After training and testing I was able to predict a sale on July, 4th 2017 of $183752.00

To a better prediction:

More data are needed

Since 4th of July is a holiday, previous holidays sales will be welcome

**6-Is there any additional information, data or access cd that would make your prediction better?**

Historical data

Add a frequency to the reported data