

IEORE4523 DATA ANALYTICS Final Project

Revenue Analytics - using Google Store under LightGBM framework as an example

December 3, 2018

Aries Li , Hung-Yu Chou, Jordan Hsieh, Wenyu Zhou

Our problem and goal - Revenue Analytics

The **80/20 rule has proven true for many businesses**—only a small percentage of customers produce most of the revenue. As such, marketing teams are challenged to make appropriate investments in promotional strategies.

We are challenged to analyze a Google Merchandise Store (also known as GStore, where Google swag is sold) customer dataset to **predict revenue per customer**. Hopefully, the outcome will be more actionable operational changes and a better use of marketing budgets for those companies who choose to use data analysis on top of GA data.

Goal for this project

Using the 3 GB dataset Google provides, we aim to:

1. Prove 80/20 rule in GStore
2. Giving recommendations based on top 3 feature importance on the revenue
3. Delivering strategies after figuring out Persona Canvas

Executive Summary

80/20 rule in GStore is confirmed

Most transaction revenue comes from a little part of customers (~1%).

Top 3 feature importance for GStore's revenue

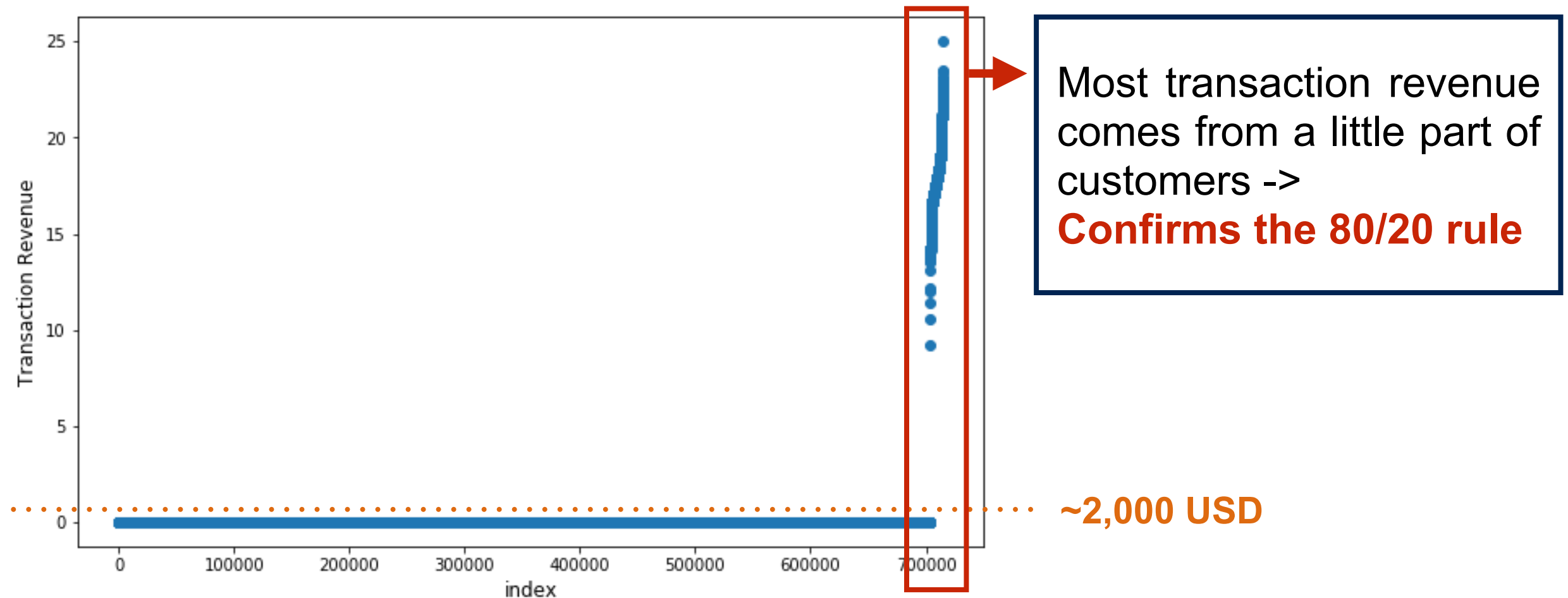
Under LightGBM framework, GStore should consider to increase customers' pageviews, visiting time, and visiting numbers on its website, **especially increase visitors' page views and hits to interval [10,20]**

Strategies for GStore by Persona Canvas

1. GStore should increase the conversion rate of its most visitors.
2. GStore should consider expand market in EU & Asia.

The 80/20 rule lies in GStore's revenue data

Transaction revenue v.s. customers



- Number of entries with none zero revenue: 11515
- Ratio between non zero entries and total entries: **0.012 (~1.2%)**
- Number of unique customers with none zero revenue: 9996
- Ratio between non zero unique customers and total entries: **0.011 (~1.1%)**

Using LightGBM

LightGBM, Light Gradient Boosting Machine

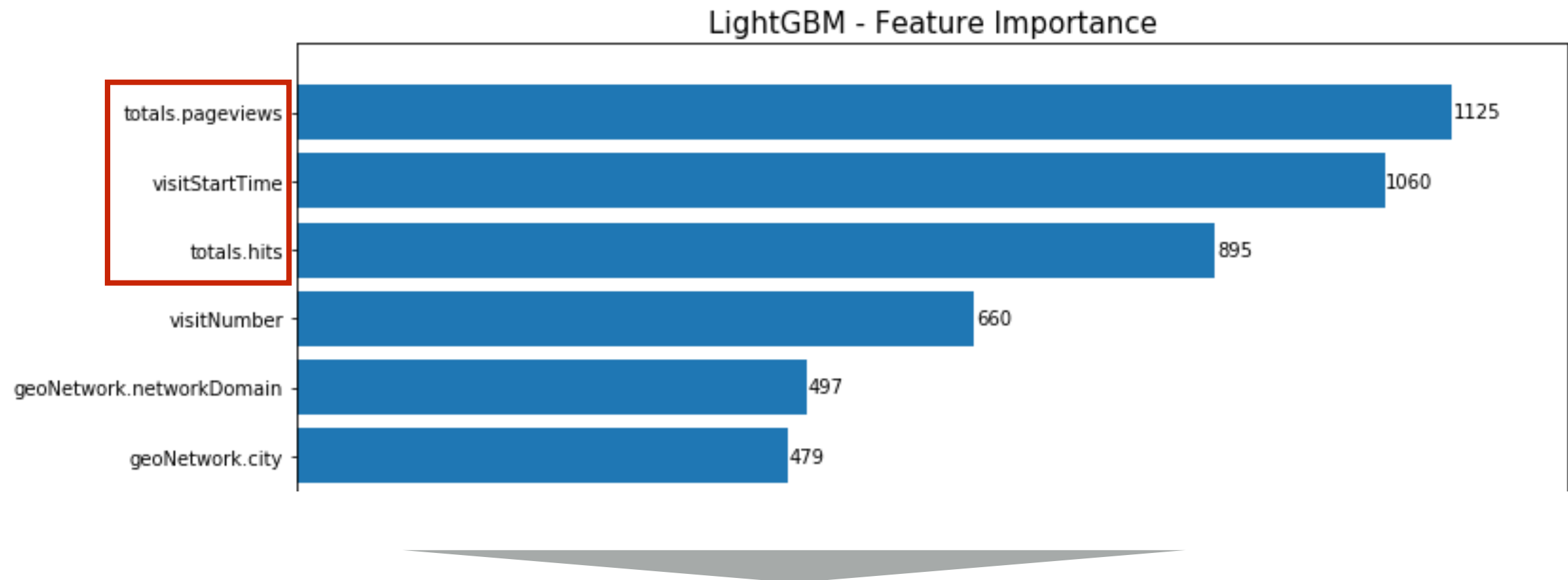
LightGBM is a gradient boosting framework that uses tree based learning algorithms. It is designed to be distributed and efficient with the following advantages:

- Faster training speed and higher efficiency.
- Lower memory usage.
- Better accuracy.
- Support of parallel and GPU learning.
- Capable of handling large-scale data.

Using LightGBM to perform ML analysis on the training dataset

3 most important factors impact transaction revenue

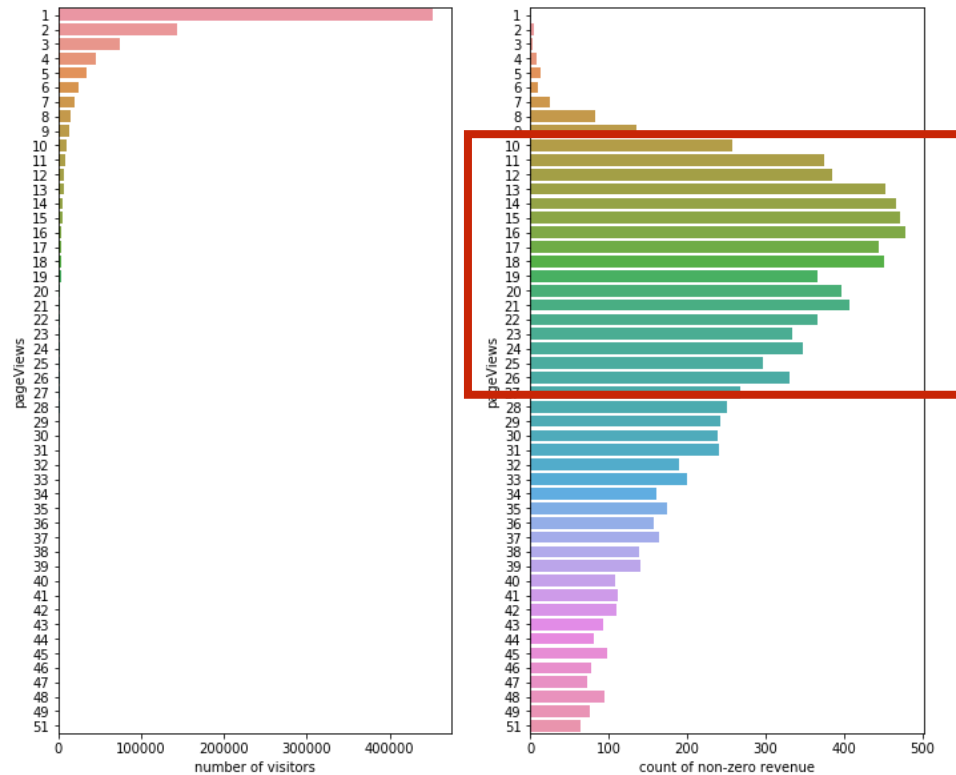
Feature importance under LightGBM



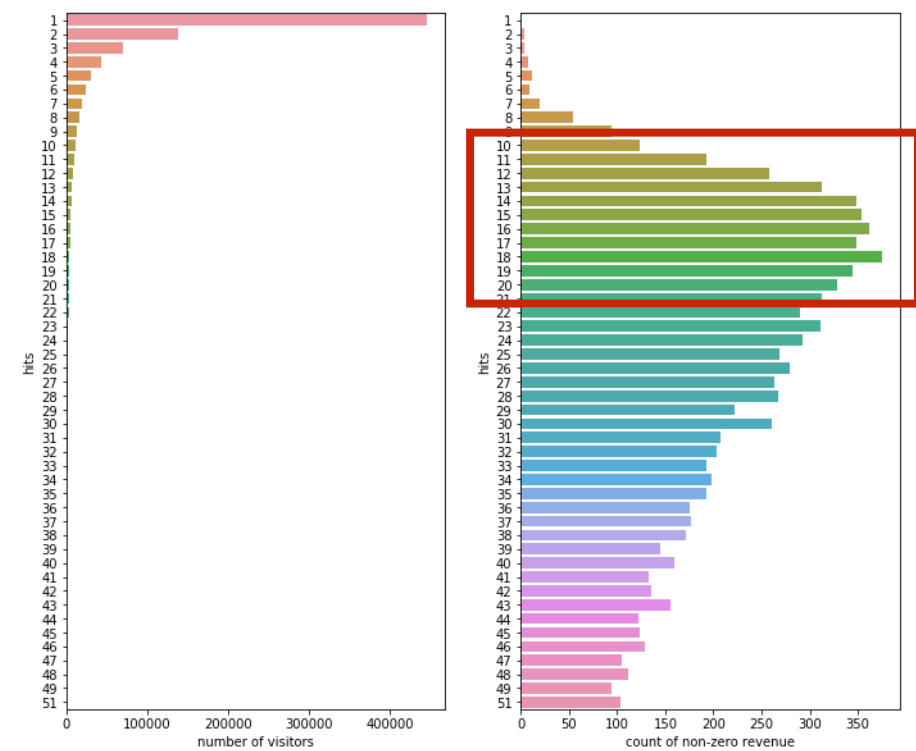
To stimulate the transaction revenue in GStore, Google should consider to increase customers' intention on browsing and hitting more pages on specific time.

Go deep into the effect of Pageviews and Hits

Page views effect on revenue



Hits effect on revenue

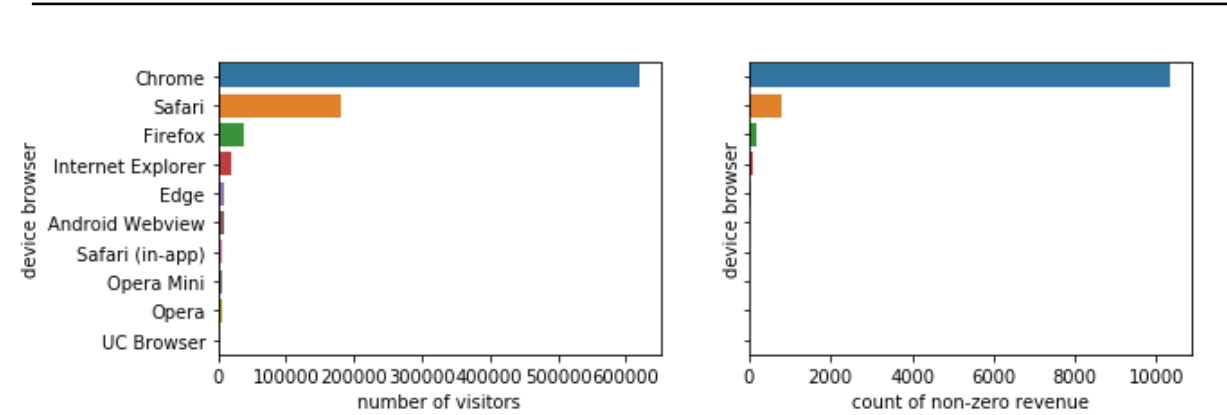


1. Most visitors have less page views and hits.
2. Number of hits and number of visitors are negative related.
3. Large amount of hits and pageviews around (10,20) has large non-zero revenue.

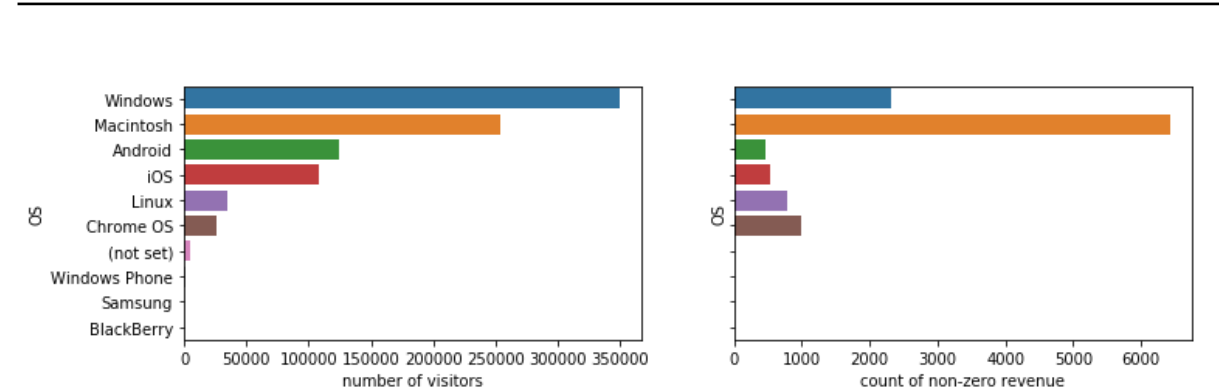
Persona Canvas for GStore

Windows Desktop using Chrome are users having most visiting counts

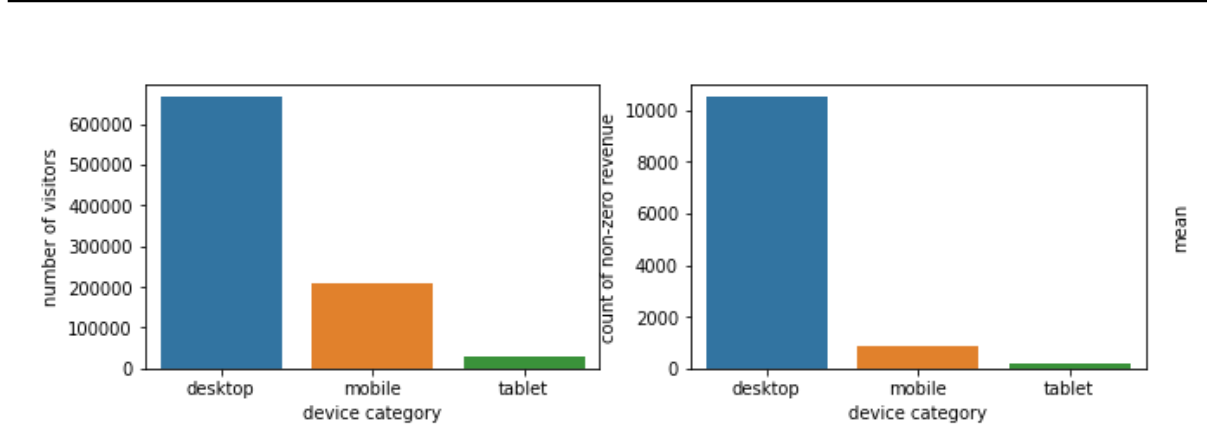
Browser



Operating System



Device Category



Mac counts the most revenue, Windows has revenue potential

Persona Canvas for GStore (cont'd)

Geographical Analysis

American

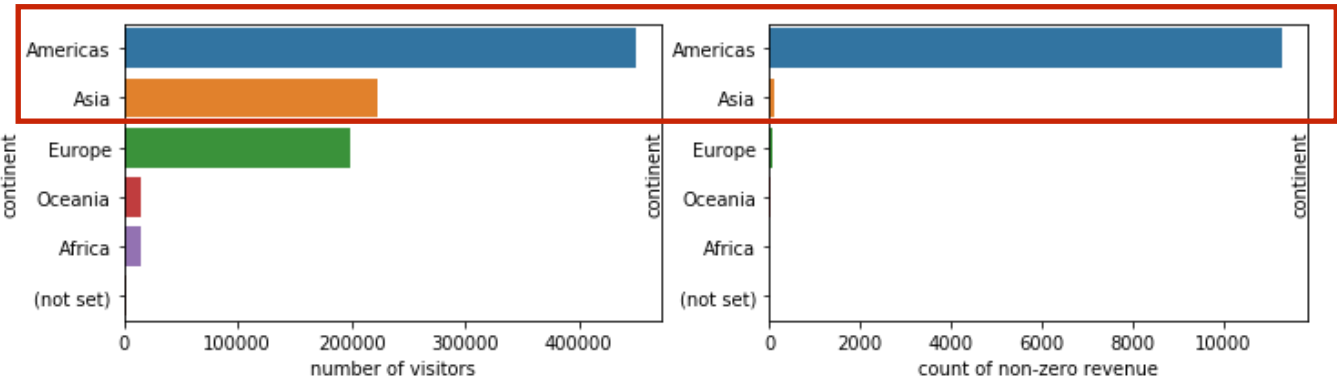
American has higher number of visitors and higher number non-zero revenue.

Asia & Europe

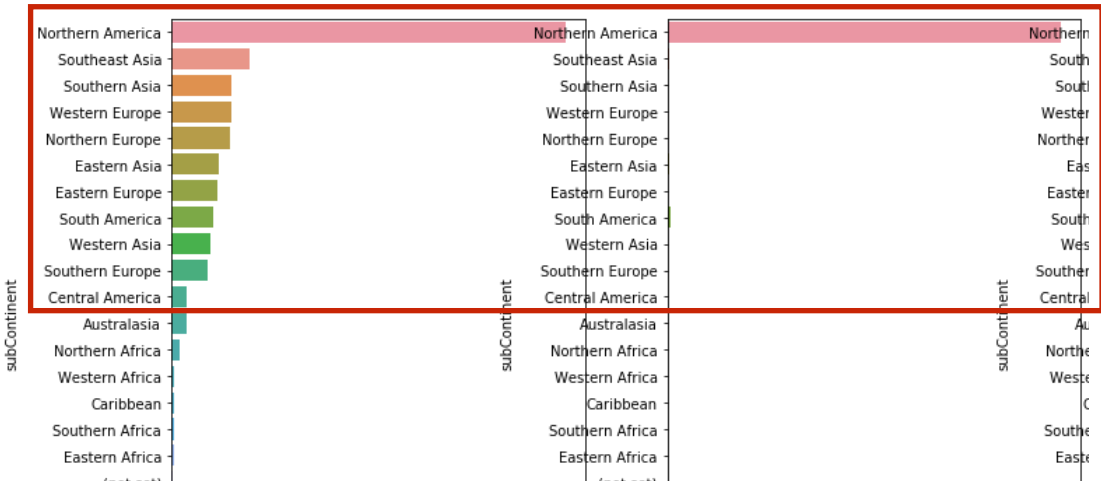
Non-trivial number of visitors, but contribute little revenue.

The above two points were further confirmed by the SubContinent plot.

Continent level - visitors and revenue contribution



SubContinent level - visitors and revenue contribution



Conclusion

80/20 rule in GStore is confirmed

Most transaction revenue comes from a little part of customers (~1%).

→ **Using 2,000 USD as the boundary to segment customers and to derive different strategies.**

Strategies for top 3 feature importances

Under LightGBM framework, customers' pageviews, hits and visiting time impact most on the revenue.

→ **Customizing its website or making promotion to target visitors' page views and hits to interval [10,20]**

Strategies after analyzing Persona Canvas

1. Windows Desktop using Chrome are users having most visiting counts, Mac users contribute the most revenue yet.

→ **Increasing the conversion rate of Win visitors.**

2. Customers in America drives more revenue.

→ **Considering expand market in EU & Asia.**