# Google Analytics 360 Case Study

Presented by:
Paul Owe

Last Updated:
May 29, 2022

# Analytics Procedure Overview


## 5. Conclusion

Recommendations

## 4. K-Means Clustering

Deriving better customer segments: Davies Bouldin index 0

## 3. Association Analysis

Advanced analytic queries to derive rule-based models

## 2. Exploratory Data Analysis

Sampling strategies, frequency distributions, correlation matrix

## 1. Introduction

Problem definition, purpose statement and proposed methodology.

# Objective

Derive “interesting” user insights from GA360 data that contribute to higher business revenue

# Problem statement: Deriving interesting user insights

In this business case

- The dataset provides **12 months (August 2016 to August 2017) of Google Analytics 360 data** for Google's merchandise e-commerce site
- The data includes Traffic source data, **Content data**, and Transactional data
- The **Content data** contains information about the **behavior of users** on the site, such as URLs of pages that visitors look at, how they interact with content, etc.
- The goal of this exploration is to **derive interesting user insights**, some of which may be our previous customers, and therefore our focus for EDA will be on the Content data within GA360 sessions.

# Exploratory Data Analysis

# EDA Tasks

The goal of this EDA is to look at the **901,097 Visits** and 1**2,070 transactions** and report high level customer information. The task will involve developing basic queries that analyze the following:

1. Total unique customers
2. Average business revenue per visitor
3. Average transactions per year
4. Buyer vs Non-buyer characterization

How many unique users visited the site in the past year?

58570 users

What is the average amount of revenue gained per session?

$120

# What is the average number of transactions per purchaser within the year?

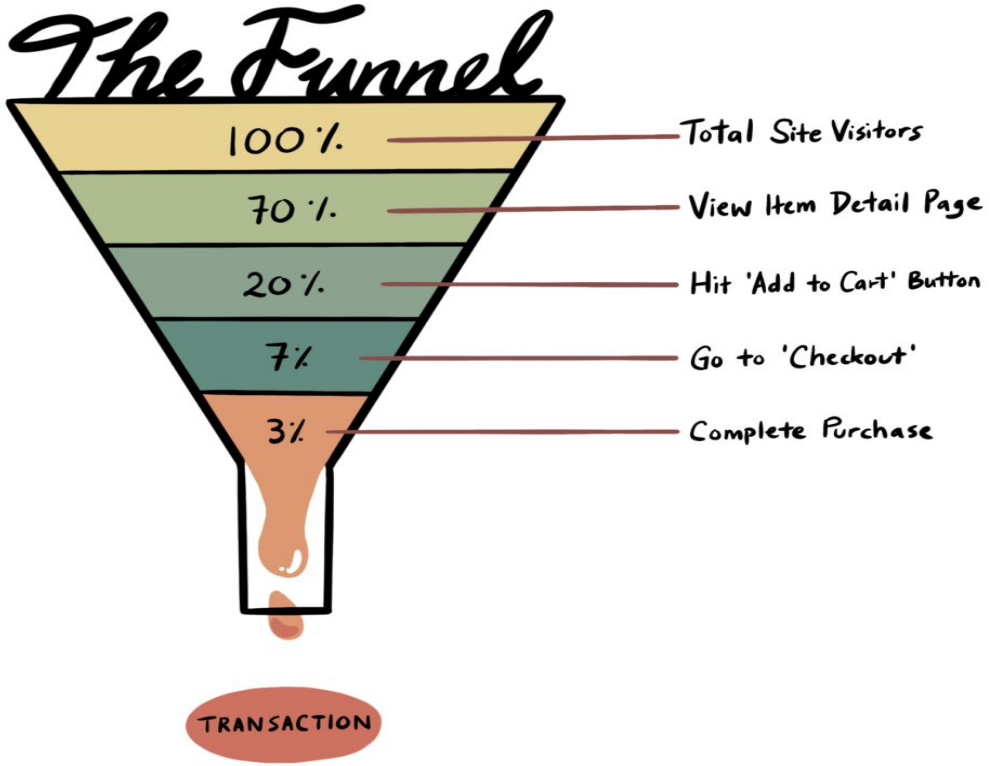
1 transaction per purchaser

# Buyers

On average these visitors had 27 product pageviews (6x more than non-buyers)

# vs Non- Buyers

On average had 4 product page views




The Funnel Illustration

# Association Analysis

Finding interesting frequent transactions in the GA360 dataset (basket data analysis)

Products purchased by a customer, who purchases eco-friendly products


Still developing insights...

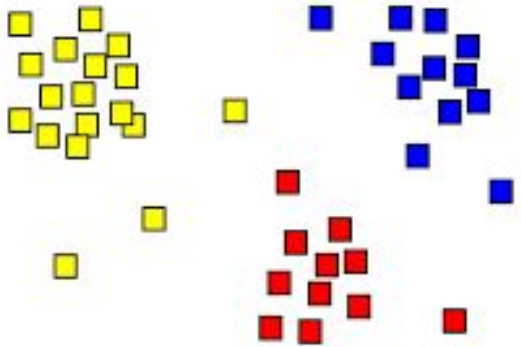
# Customer Segmentation: K-Means Clustering

# Building a k-means clustering model for market segmentation

**Objectives**

1. Determining themes in your consumer behavior, that can be **used to develop customer profiles** or **personas**
2. Explore features to understand what might be interesting for a clustering model
3. **Personalize your Ad campaigns** to customers by tailoring ads and other communications to them

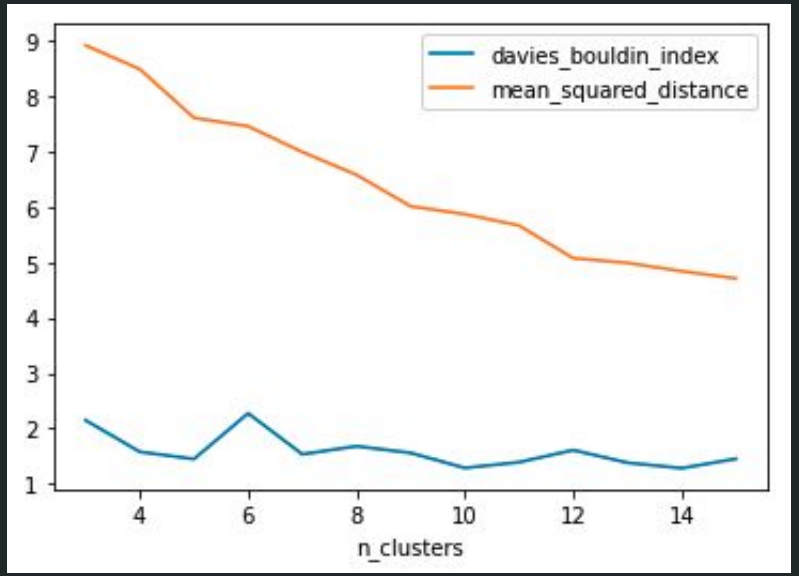
Clustering with high intra-class similarity, and low inter-class similarity



# Cluster validity

- With 5 clusters, we see a score of ~1.4, and only with k>9, do we see better values wrt DB index.

Selecting k clusters which minimize DB index value



# Cluster interpretation (k=5)

Cluster 1 - The apparel shopper, which also purchases more often than normal. This (although synthetic data) segment skews female.

Cluster 2 - Most likely to shop by brand, and interested in bags. This segment has fewer purchases on average than the first cluster, however, this is the highest value customer.

Cluster 3 - The most populated cluster, this one has a small amount of purchases and spends less on average. This segment is the one time buyer, rather than the brand loyalist.

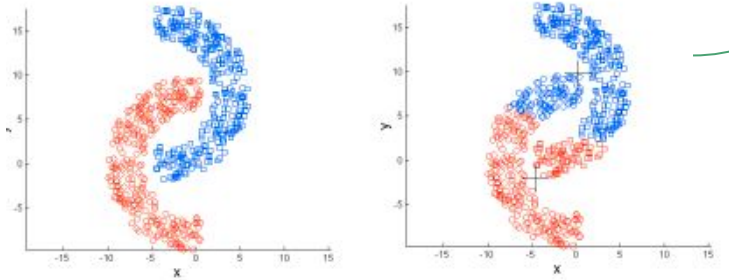
Cluster 4 - Most interested in accessories, does not buy as often as cluster 1 and 2, however buys more than cluster 3.

Cluster 5 - This is an outlier as only 1 person belongs to this group.

# Comments on the K-Means Model

**Strengths**

1. Our model gains more insight into **a holistic view of our customers**
2. Relatively efficient O(tkn), where n is the number of objects, k is the number of clusters and t is the number of iterations



**Weakness**

1. Sensitive to initial clusters
2. Need to specify K
3. Sensitive to noise and outliers
4. Not suitable for non-convex shapes

**Alternative**

- K-Medoids (Partition around medoids) is more robust

# Business Recommendations

# Key takeaways and recommendations

**What generates revenue?**

- Sales from the google brands, and apparels
- Business should put in place strategies to build brand loyalty
- A mix of sales, developing leads for sales, and increasing page views are good indicators

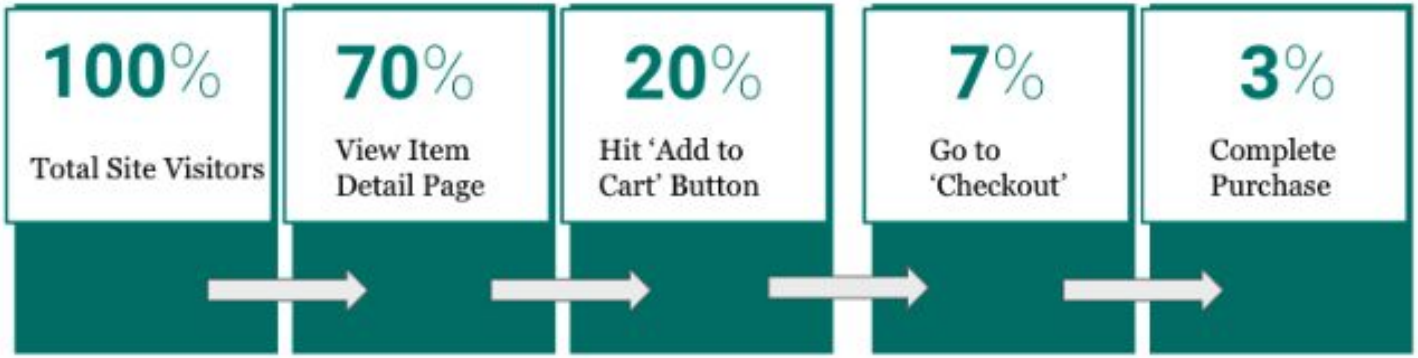
**What to do with observed frequent patterns?**

- Place items frequently bought together close to each other to increase sales

**Customer segmentation**

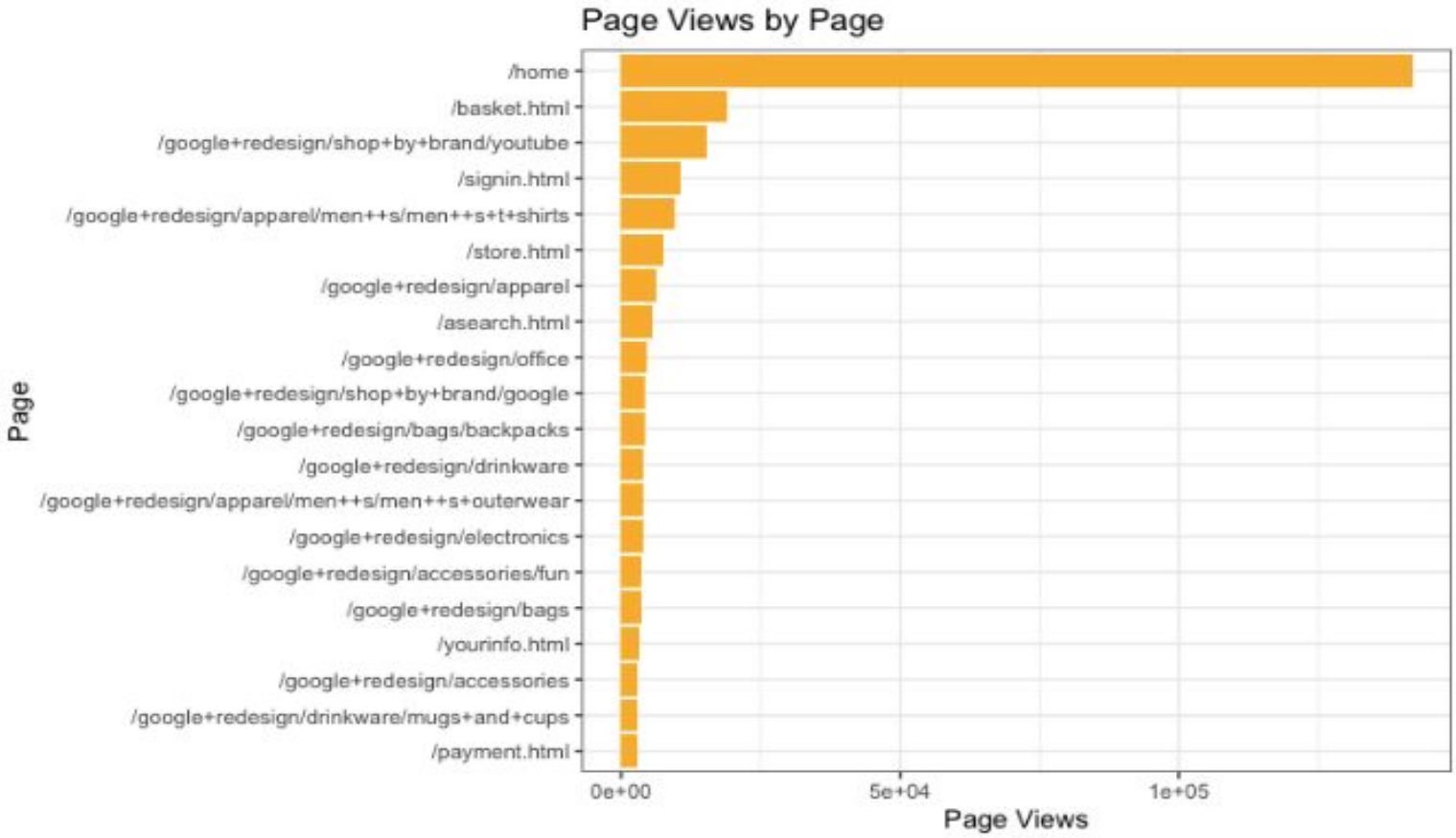
- One time buyers
- Females have a tendency of purchasing more often than men
- Customers prefer shopping by brand, and this segment is the highest value

# Visualizations artifacts

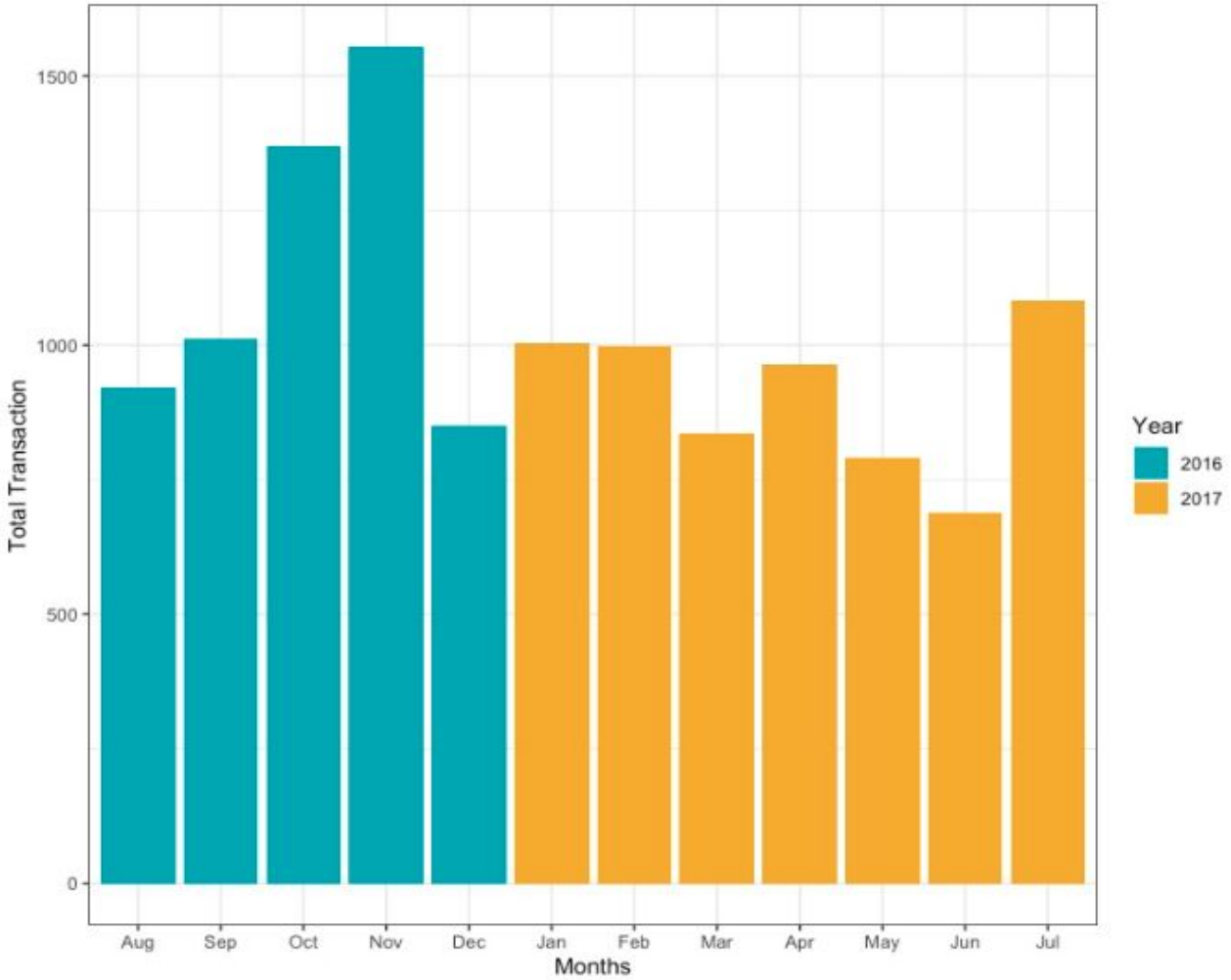



| | | | | |
|---|---|---|---|---|
| Marketing | # Items Shown | Product description | Calls to action | Card on file |
| SEO | Price | # of images | Promotions | # of fields to complete |
| Seasonality | UX Quality | Ratings and reviews | | Registered user vs guest |

# Popular pages by pageviews


Page Views by Page
/home
/basket.html
/google+redesign/shop+by+brand/youtube
/signin.html
/google+redesign/apparel/men++s/men++s+t+shirts
/store.html
/google+redesign/apparel
/asearch.html
/google+redesign/office
/google+redesign/shop+by+brand/google
/google+redesign/bags/backpacks
/google+redesign/drinkware
/google+redesign/apparel/men++s/men++s+outerwear
/google+redesign/electronics
/google+redesign/accessories/fun
/google+redesign/bags
/yourinfo.html
/google+redesign/accessories
/google+redesign/drinkware/mugs+and+cups
/payment.html
Page
0e+00
5e+04
1e+05
Page Views

# Total transactions


1500
1000
500
0
Total Transaction
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Months
Year
2016
2017

# References

# References


BigQuery Cookbook

Customer Segmentation with Python

K-means clustering models for market segmentation using BigQuery ML

Web analytics on Google online merch store

# Thank you