Google Analytics Customer Purchase Prediction

Yuji Mori

Stats 404
Winter 2021

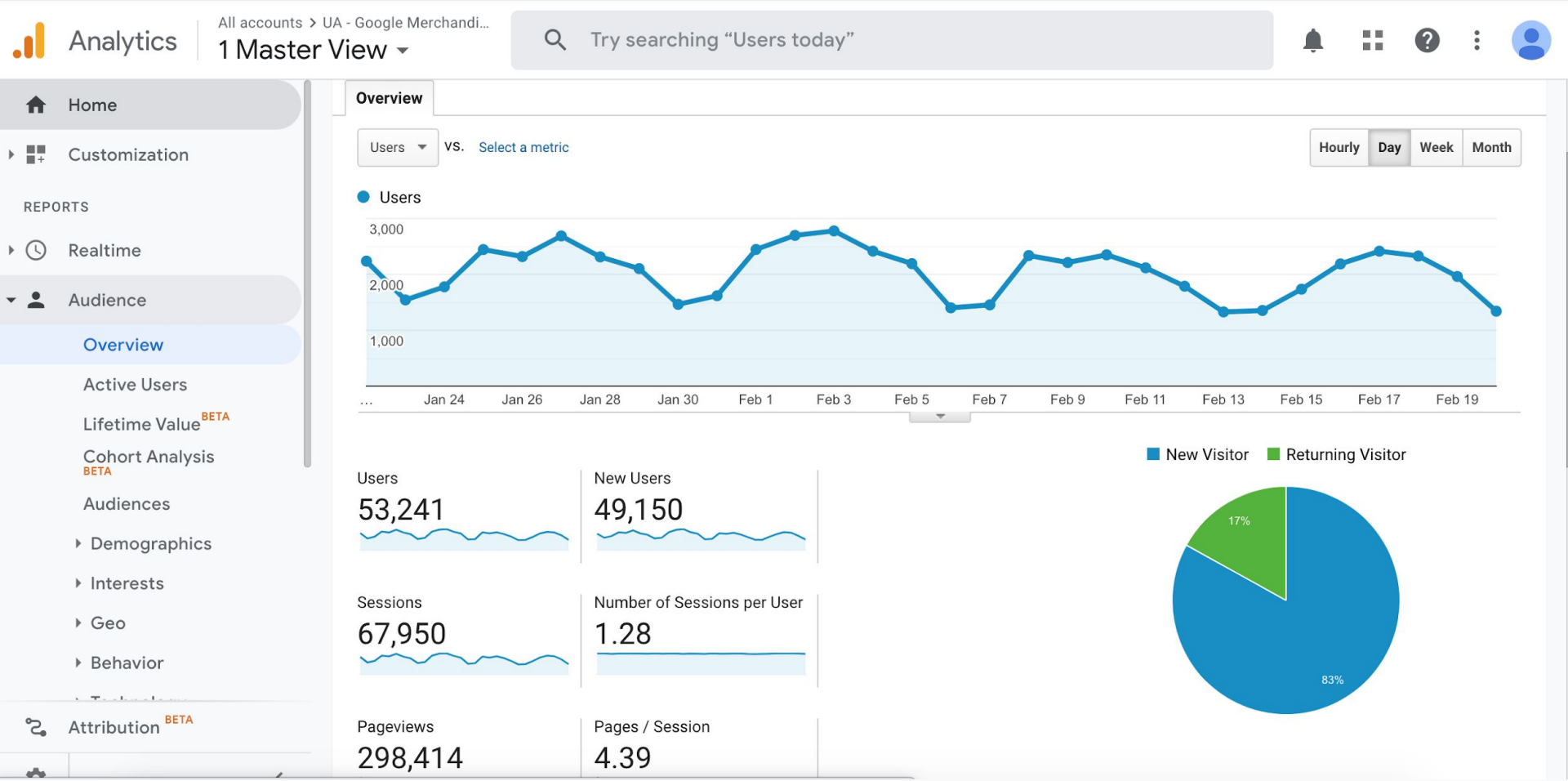
Introduction - Google Analytics

Google Analytics *is a web analytics service that tracks/reports website traffic data.*

Commonly tracked metrics include:

- **Demographics** (Age, Gender) and **Geographic data** (Language, Location)
- **Web Engagement** (time spent on site, # of pages clicked, etc.)
- **Acquisition** (how did visitors find the website? Did they click on an ad?)
- **Conversions** (a success event defined by the business, such as a store purchase)

These metrics are used to strategically improve marketing/advertising!



(screenshot of Google Analytics Dashboard, using Demo Account)

The Business Problem

- **Suppose an online store/business uses Google Analytics ...**
  - They have general idea of their audience and sales trends.
  - Example: 25-35 year old males from U.S. are their top purchasers
  - Most visitors in this group buy nothing!

Key Business Question:

Can we leverage all the features of a given web visitor to better predict whether they will make a purchase?

The Data - Google Merchandise Store

- From Kaggle competition: Google Analytics Customer Revenue Prediction
  - The challenge: *analyze a Google Merchandise Store customer dataset to predict revenue per customer*
  - Adapted to predict event of purchase (classification problem)
- Mainly Geographic and Acquisition data (no personal info)
- Data is in a standardized structure (via **BigQuery** , Google Cloud big data platform),
  - **ANY business using Google Analytics can leverage predictive models!**

Business Impact of Work

- Provide an “out-of-box” solution for businesses to leverage ML on their Google Analytics data
  - Determine whether any given web visitor will make purchase
  - Automatically output top model predictors for revenue
  - No need to struggle with Google Analytics web platform
- Allow business to focus marketing/advertising efforts based on these predictions

Business Impact of Work

Example:

Suppose model insights lead to **20% increase** in conversion rate (1.5% → 1.8%)

- Monthly Revenue *: **$52,500** ⟶ **$63,000**

+ $10,500/mo. → **Over $125,000 additional revenue generated per year!**

* Assuming: ($50) average revenue per sale, (70,000) monthly visits

Formula: (AVG $ per sale) x (monthly visits) x (% of visits with sale) = Monthly Revenue

Methodology

- Implemented a **Random Forest Classification Model** to predict event of purchase for each web visitor
  - The algorithm creates decision rules on a visitor's features
  - Example:
    - (Country = "U.S.") && (Age > 25) ⟶ **Purchase!**
    - (Country = "Germany") && (Age > 60) ⟶ **No Purchase!**
- Model also identifies the top predictors for revenue

Model Performance

Baseline: **Targeting only U.S. visitors**

- **~2.9%** of visitors make a purchase

If 10,000 visitors clicked on ad:

- $1 Cost-Per-Click → ~ $9,700 spent
- $50 Avg. Sale → ~ $15,000 revenue

Random Forest Model:

- **~4.5%** of visitors make a purchase

If 10,000 visitors clicked on ad:

- **+20% Conversion** w/ new strategy
- $1 Cost-Per-Click → ~$9,500 spent
- $50 Avg. Sale → ~ $27,000 revenue

Overall Benefit:

+ $12,500 per 10,000 visitors!

Model Applications

- Use top model features to select target audience in future marketing + advertising campaigns
  - Google Analytics links to Google Ads service!
  - Value added by conversion % , budget savings, etc.
- Visitor ID can be placed directly into targeted lists
- Potential to stream data directly from Google Cloud for live predictions and dynamic advertising strategies

	Top Features from GStore Model
1	Continent (North America)
2	Country (United States)
3	Referral Source (mall.googleplex.com)
4	Referral Source (youtube.com)
5	Referral Channel (“Social”)
	...

Thank you for your Questions/Feedback!