Predictive Modeling for Marketing Effectiveness with Multiple Linear Regression

David Harvell
Western Governors University

Introduction - David Harvell

Education

- Oracle Certified SQL Developer
- Udacity Data Analyst Certification
- CompTIA Project+
- B.S. in Data Management/Data Analytics
- Pursuing M.S. in Data Analytics

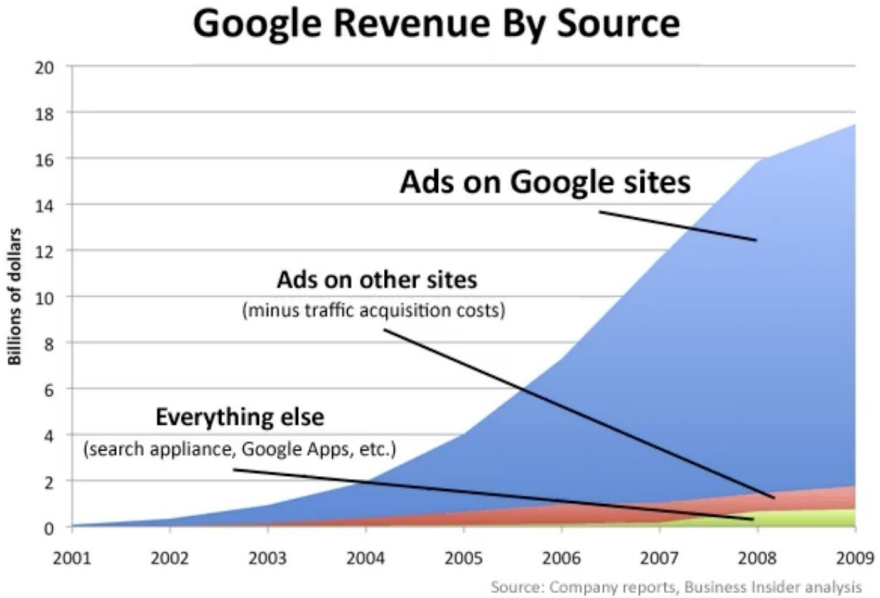
Professional

- SQL Developer for 15 years
- Lead Business Intelligence at Multiple Companies
- Retail Analytics for Past 6 Years, Creating CPG Reporting for Walmart Accounts
- Current Product Owner for Retail Analytics Software

Problem Statement and Hypothesis

How much does the frequency of ads impact the probability of purchase?

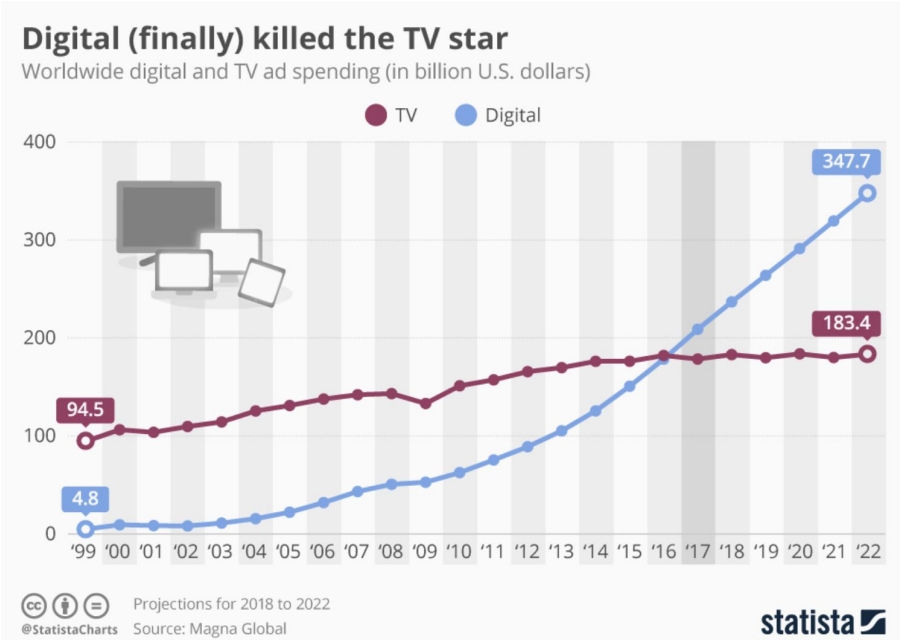
Our Problem Statement



Alphabet (Google's Parent Co.) is the 9th largest business in the word

Source: https://fortune.com/fortune500/

Online/Digital is now the dominant medium for advertising



Source: https://www.statista.com/chart/12136/worldwide-digital-and-tv-ad-spending/

NULL Hypothesis:

Total ad count has NO impact on the probability of conversion

Data Analysis, Limitations, and Outline of the Findings

Marketing A/B Dataset from Kaggle

- CSV Format
- 22 MB total size
- 588K Records
- Users shown Ads or PSAs
- Clean Dataset
- Public License

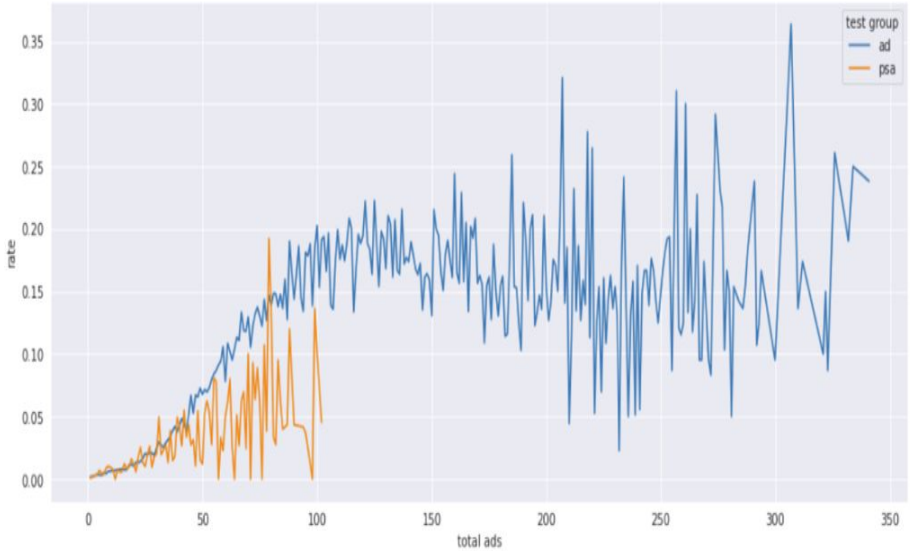
Data Dictionary

- ID Columns
- Shown Ads or PSAs
- Did User Purchase (Conversion)
- Total Ads Shown
- Day with Most Ads
- Hour with Most Ads

Source: https://www.kaggle.com/faviovaz/marketing-ab-testing

Steps to Analysis

- Visual Inspections
- Drop ID Columns
- New Rollups for Rate
- Convert Text to Numbers
- Check Correlation
- Create Multiple Linear Regression Model
- Evaluate
- Tune Dataset and Re-Model
- Evaluate



Limitations

Python

- Requires Programming Knowledge
- Slower Execution for Larger Dataset

Google Colab

- Cost Increases for Larger Dataset and More Processing

MLR (Multiple Linear Regression)

- Possible Misinterpretation of Results

4.7%

Only 4.7% of results and can be explained consistently by the dataset

Total Ads Shown IS statistically significant to conversions

OLS Regression Results

Dep. Variable:	converted_d	**R-squared:**	0.047
Model:	OLS	**Adj. R-squared:**	0.047
Method:	Least Squares	**F-statistic:**	1.462e+04
Date:	Thu, 06 Jan 2022	**Prob (F-statistic):**	0.00
Time:	12:34:54	**Log-Likelihood:**	2.6923e+05
No. Observations:	588101	**AIC:**	-5.384e+05
Df Residuals:	588098	**BIC:**	-5.384e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0007	0.001	-0.679	0.497	-0.003	0.001
total ads	1.6101	0.009	170.826	0.000	1.592	1.629
group_d	0.0076	0.001	7.503	0.000	0.006	0.010

Proposed Actions and Expected Benefits

Proposed Actions

1. Collect More Data

- Increased record count
- Increased variable/property count
- Integrate new datasets

2. Allocate Money to Digital Advertising

Money allocated to online advertising should result in an increased number of sales

Thank You!

David Harvell
Master of Science, Data Analytics Student
dharv39@wgu.edu
https://www.linkedin.com/in/davidharvell/