FINAL PROJECT REPORT
SURVEY ANALYSIS

OMNI-CHANNEL MARKETING STRATEGY

Dhanika Sujan
Koushik Subramani Murali
Shreya Bedi
Sudheeshna Sampath

ABOUT US

For our final project, we have analyzed the Deloitte Digital Democracy Survey. We call our project “Omni-Channel Marketing Strategy” as we want our customers to have a seamless experience. We want to acknowledge the previous touch points along the customer journey.

Our team is a Video Streaming Service that shows advertisements across Television, Computer and other portable media platforms and channels.

THE GOAL

Our team aims to launch a premium service with additional content based on personalized recommendations and demonstrate higher value to our customers.

ACTION PLAN

Firstly, we identified customer segments based on their sentiment towards marketing and personalization.

Then we discovered the right channel and style to market to these customer segments.

DATA WRANGLING

- Removed columns that were irrelevant to our analysis such as children per household, interest in live events

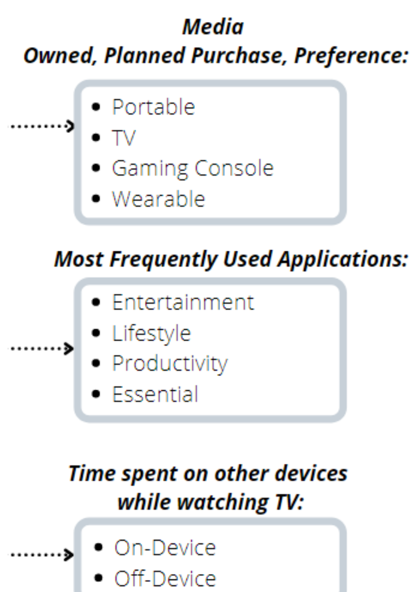
- Grouped columns such as media owned, planned to purchase, media preference by aggregating columns that fall under the following buckets: ‘portable devices’, ‘TV’, ‘smart wearables’, ‘gaming devices’ to give us the ‘Top Media Owned’, ‘Media planned to purchase’, ‘Preferred Media’

- We grouped people’s preference to the above-mentioned buckets (‘portable devices’, ‘TV’, ‘smart wearables’, ‘gaming devices’) for each form of entertainment (Movie, TV shows, Sports) to get ‘Device for Movie’, ‘Device for Sports’, ‘Device for TV shows’ columns

- Most frequently used apps such as ‘Business’, ‘Streaming music’, ‘Weather’, ‘Browser’ etc were grouped into 4 major buckets: ‘Essentials’, ‘Entertainment’, ‘Productivity’, ‘Lifestyle’. Taking the total of the apps in each bucket and picking the bucket with the maximum ‘yes’ we get the ‘Most Frequent Apps Used’ column.

- We then grouped the attached media for users while watching home television to 4 buckets: ‘On-Device’ and ‘Off-device’ based on if the activity was something that involved an electronic device or not. We then took the Percent of time spent on-device to get the ‘Time Spent on Device while watching’ column.

- Based on user’s reaction to advertisement i.e., if they would rather pay more to get no ads, being ok with ads if it means a subscription discount, think data sharing leads to identity theft, being ok with ads if it is targeted, we group them and created the ‘Market affinity’ column to see if they would respond favorably or not to ads.



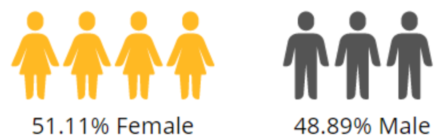
EXPLORATORY DATA ANALYSIS

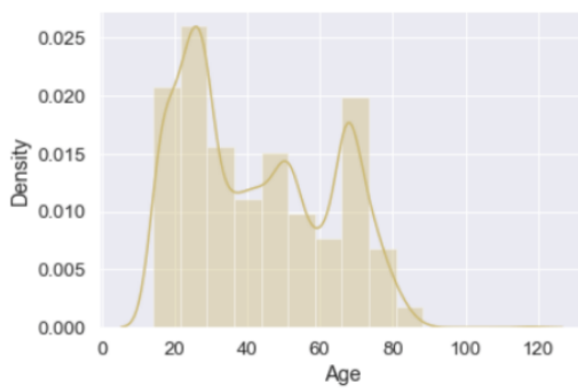
Following are the major analysis that we wanted to conduct on the data to come up with various strategies in order to achieve our aim:

Age and Gender Distribution Analysis:

Looking at the charts below, we can see that the female population is a little more compared to the male population who took the survey.

<u>Distplot for Age Distribution using kernel density estimate (KDE):</u>
From the plot we can see that the survey taking population is majorly in their 20s to 40s. We can also see a sudden spike around 65 to 70 years of age.





Device Distribution:

We grouped people's preference ('portable devices', 'TV', 'smart wearables', 'gaming devices') for each form of entertainment (Movie, TV shows, Sports) to get 'Device for Movie', 'Device for Sports', 'Device for TV shows' columns.

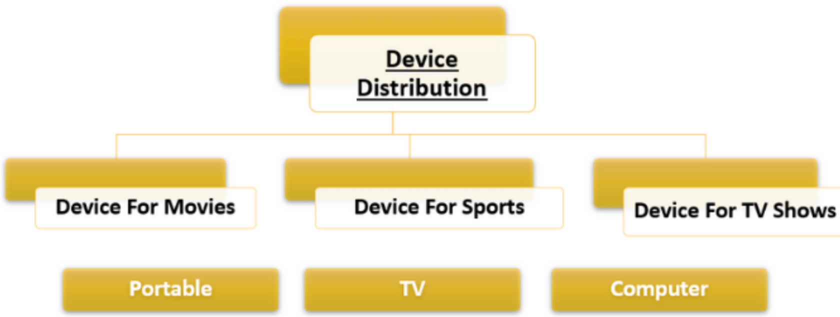
Devices are distributed in the following categories:

a. Devices for movies
b. Devices for Sports
c. Devices for TV Shows

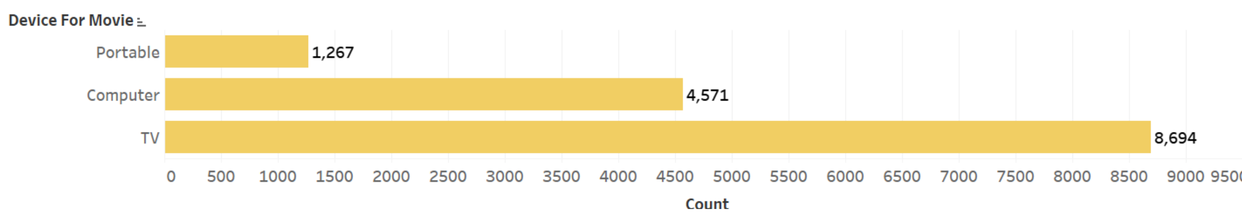
Which are further divided in the following categories:

i. Portable

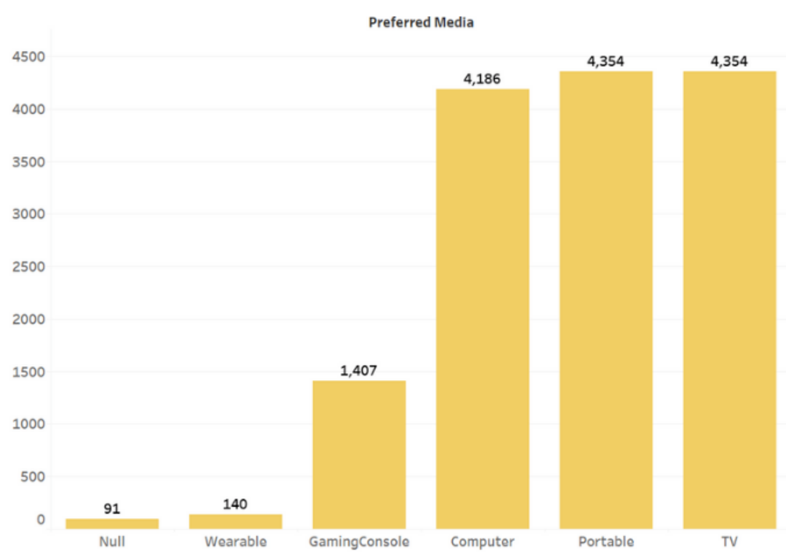
ii. TV
iii. Computer



From our analysis, we could see that "TV" was the most popular form of device used by the customers for all the above mentioned categories. Below is a sample chart created to show one of the examples of how usage of these devices is distributed:

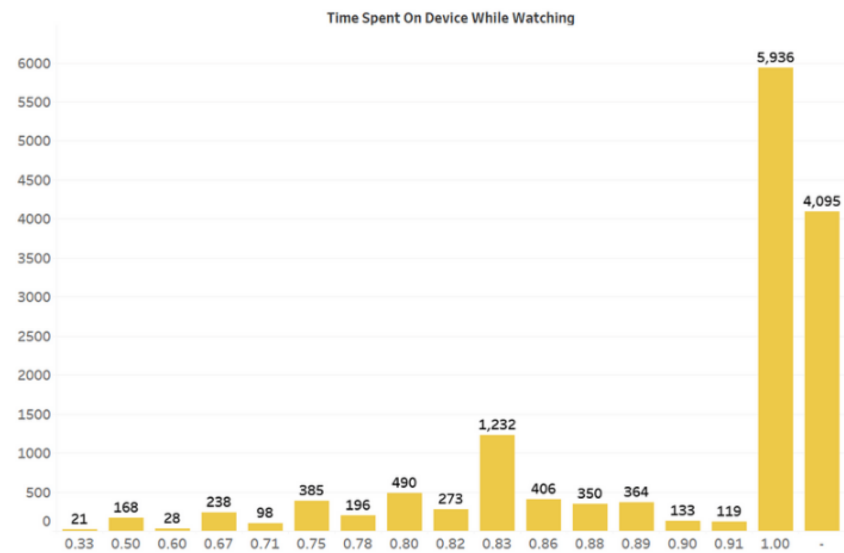


Preferred Media:



Looking at the graph, we saw that there are a few records where we do not have preferred media information. Apart from that, customers usually prefer using Television or other portable devices.

Time Spent On Device While Watching TV:



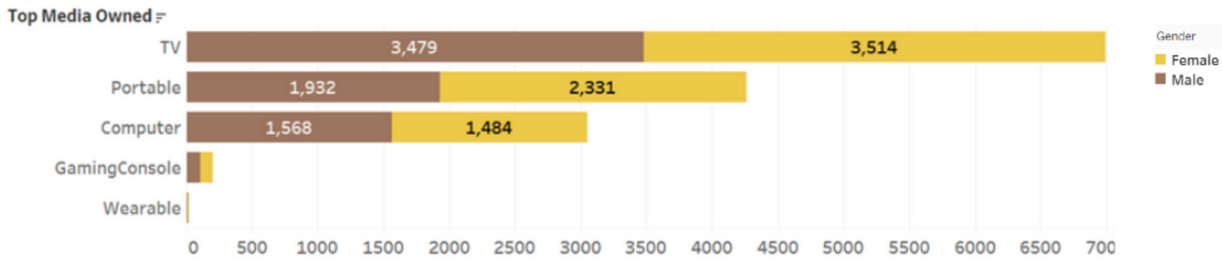
There are a lot of null values for understanding how much time a user spends on other devices while watching television. We can also see that many spend almost all of their time on other devices while watching television.

Top Media Count:

Top Media Owned	
TV	6,993
Portable	4,263
Computer	3,052
GamingConsole	203
Wearable	21

As we have discussed above, TV and Portable devices are the preferred form of devices owned by the users.

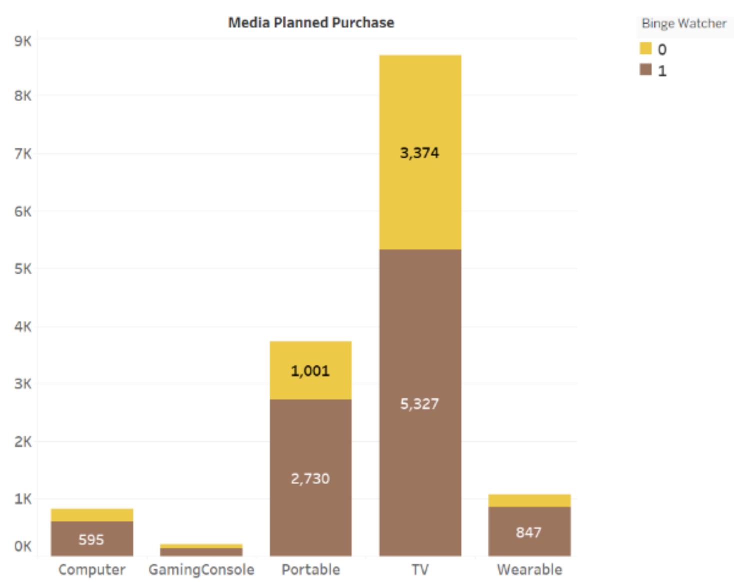
Top Media (Distributed by gender):



Television is the top most media preference for both male and females. They are almost equally divided when comparing the count distribution.

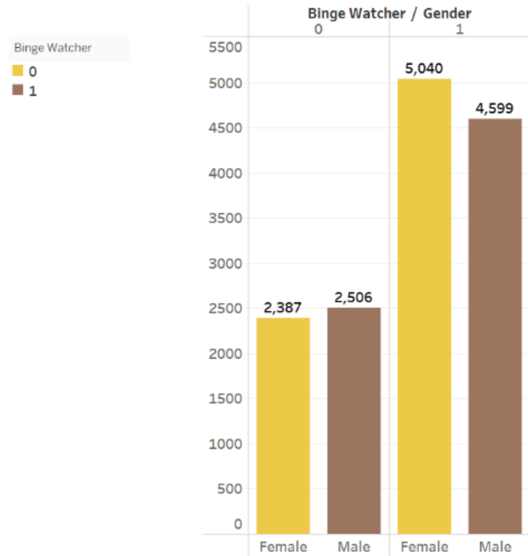
After television, other portable media devices are preferred by the users where the female population dominates in owning these devices.

Binge Watchers (Distributed by media):



Looking at the graph, we can see that Television is definitely the most popular media platform used by the users to binge watch. Apart from television, other Portable devices are famously used to binge watch. This is useful information for us to come up with our plan and strategies.

Binge Watchers (Distributed by gender):



Looking at the above graph, we can see that we have more users who are binge watchers compared to non-binge watchers. In the binge watch category, the female population dominates over the male population.

Apart from the above discussed exploratory data analysis, we conducted a few other analyses as well. Below are the mentioned pointers for the same:

- Analyzing the type of applications used, entertainment and productivity type of applications dominated.
- Major survey taking population belongs to medium to low income category. We also have a few records where this data is not available.
- Major chunk of the survey taking population is currently employed.
- Survey taking population is majorly from the South or the West region of the country.

CUSTOMER SENTIMENT TOWARDS MARKETING

In order to better understand how our customers felt about marketing strategies and advertisements in particular, we came up with the Market Affinity feature. This feature will help build our recommendation model and provide targeted recommendations regarding advertising strategies for each of the customer segments. The Market Affinity feature groups the customers into the below 4 categories:

1. **High Interest Customers**: These customers are extremely interested in targeted advertising and are actually even willing to pay higher prices for the premium subscription in exchange for personalized recommendations and targeted advertisements. These would ideally be our premium customers who would benefit the most from the premium service.

2. **Neutral Customers**: These customers are not really interested or expect targeted advertisements from the service. They do however expect subscription discounts and perks if they are shown advertisements.

3. **Disinterested Customers**: They do not wish to see advertisements, be it general or targetted. These customers are even willing to pay a higher price just so that they can avoid advertisements. We would have to come up with tailored recommendations to make sure the disinterested customer sentiments are taken into account in our new product.

4. **Unfavorable**: Customers under this group have strong opinions about advertisements and are completely unwilling to share any personal information for targeted advertisements or recommendations. They view advertising as a medium for privacy threats and identity theft.

DATA PRE-PROCESSING STEPS

The initial dataset had 1800 records spread across 15 features/columns and consisted of continuous, nominal and ordinal variables. We carried out the below pre-processing steps to get the training data ready for modeling:

1. **Label Encoding:** We used sklearns LabelEncoder to normalize the data and convert the target variable into numeric values so that it is ready for modeling. After encoding the target, there were 1800 records and 41 columns in the training data.

2. **Oversampling:** The data available was highly imbalanced with unequal representation of all customer segments. RandomOverSampler was used to fix this by oversampling the minority groups by choosing records for random replacement. This step increased the number of rows in the dataset from 1800 to 3500 records. The final dataset has about 3500 records, with 15 features that are spread out across 41 columns.

MODEL SELECTION AND EVALUATION

We tried to fit 3 different classification models on our dataset inorder to predict customer sentiment. The dataset was split into 80% train and 20% test data. The target variable we are trying to predict is the MarketAffinity feature as described earlier. We also performed Hyperparameter Tuning while developing the model.

1. **Random Forest** : RandomForestClassifier was applied on the training dataset to predict Market Affinity after using cross-validated grid search to identify the optimized hyper parameters. The hyperparameters selected after grid search are as below.
   max_depth: 12
   max_features: 'sqrt'
   min_samples_leaf': 1
   min_samples_split: 2
   n_estimators: 250

2. **Decision Tree :** This supervised learning model helps predict the market affinity by learning from the different features we have shortlisted in our analysis. The DecisionTreeClassifier using the below hyperparameters does not perform as well as Random Forest. A decision tree with a maximum depth of 12 gives the best results.
   criterion: ‘gini'
   max_depth: 12

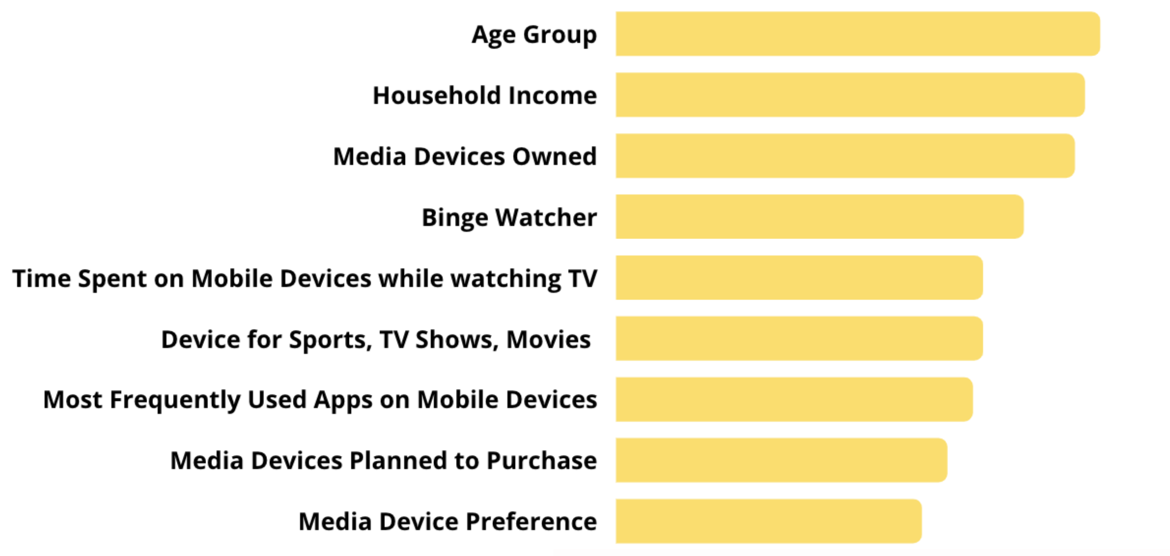
3. **XGBoost :** The gradient boosted decision tree model was applied to see if there was any increment in the speed and performance of our model. This supervised model to predict the target again did not perform as well as the initial Random Forest Model.

Three main metrics were used to evaluate each of the model performance and choose the best model. The metrics are : The overall accuracy of the model, the accuracy of the predictions on the test data and the F1 Score. The below table summarizes the findings.

Model	Overall Accuracy	F1 Score	Test Data Accuracy
Random Forest	90%	90%	71%
Decision Tree	78%	77%	61%
XGBoost	73%	71%	58%

Random Forest was the best model according to our evaluations and this is the model we are choosing to base our recommendations on.

As part of the evaluation, we discovered the key features which are the most useful in predicting our target Market Affinity variable.



We also created a correlation matrix to weigh the importance of each of these features with the target Market Affinity variable and further performed data analysis on these features to see if we could extract any further patterns and insights for our recommendations.

CUSTOMER SEGMENT LEVEL FINDINGS AND RECOMMENDATIONS

Based on the model selection and evaluation, we have listed our findings and formulated the below recommendations for each of our 4 customer segments. The premium service we are proposing will be built surrounding these findings and recommendations. We believe these recommendations will help build a premium service which meets the different customer segments advertising requirements and provides top customer value.

Overall we see that 46% of the surveyed population do not welcome targeted advertisements.

1. **High Interest Customers:** Majority of these customers are millennials and GenZ and consume their media on television and portable devices. A lot of their television watch time is spent consuming some form of content on other devices. Since these consumers spend a lot of time on social media and entertainment applications, we recommend that marketing surveys are sent to these customers over social media. The results will help base our premium service's personalized recommendations and advertisements for them.

2. **Neutral Customers:** These consumers are again majorly millennials and GenZ and have lower incomes (probably as they are students) . They primarily binge watch their media content. Since they comprise the younger population who are open to ads if they pay a lower price, our strategy to this segment would be to provide student discounts or various offers on completing the marketing surveys over social media.

3. **Disinterested Customers:** These are millennials and GenZ who spend a lot of time on portable devices for entertainment, social-media and occasionally binge on shows and consume media on television. Since they have a high willingness to pay for ad-free premium services, our strategy is to provide personalized content recommendations and banner ads on social media or television.

4. **Unfavorable:** These are the Boomers & Gen-X population who mostly only use their mobile phones for essentials and productivity tasks. They own and consume media through television and the computer and are not active users on social media. Our primary marketing strategy for this group is to have awareness based advertisements on the television.

ADDITIONAL NOTES

Everyone had roughly equal contributions to the project. In particular, the data processing and EDA was largely handled by Shreya Bedi and Dhanika Sujan. Koushik Subramani Murali and Sudheeshna Sampath contributed largely to the different model selection and evaluation part.

REFERENCES

- Images sourced from Google
- Data sourced from Deloitte Digital Democracy case study