Deloitte Survey Analysis

HyperNet — a Home Internet Service Provider

Ruchi Bhatia, Caroline Zhu, Maxine Ma

1. Executive Summary

HyperNet is a home internet service provider that has been serving customers for several years. As the demand for faster internet speeds continues to grow, HyperNet is considering offering an internet speed upgrade option to its customers.

As a team of Data Analysts working at HyperNet, we are committed to providing insights into why customers demand faster internet speeds and identifying patterns in customer demand. To do this, we will utilize the Deloitte Media Consumption dataset. Our analysis aims to inform strategic decision-making for the company, helping it remain competitive in the home internet service provider market.

Our analysis using Logistic Regression and Clustering techniques like K-Means and Agglomerative clustering has revealed significant insights into customer behavior and preferences regarding internet usage. Specifically, the frequency of app usage and age were identified as the most important factors influencing willingness to pay for higher internet speeds. Additionally, the Clustering analysis identified distinct customer segments based on their internet usage behavior, providing valuable information for targeted marketing and service development. By leveraging these insights, internet service providers can better understand their customers and develop tailored strategies to improve customer engagement and satisfaction.

Based on the modeling result, further feature analysis revealed that individuals willing to pay exhibit higher usage frequency of social network platforms, streaming music services, and mobile games, compared to those who are unwilling to pay. Moreover, they spend more time watching TV shows and movies on their tablets than those who are unwilling to pay. The target personas of customers who exhibit willingness to pay are young adults who are predominantly employed full-time or part-time, and families falling within the annual income range of 50,000 to 99,999 dollars. Regional and state-wise distribution analysis revealed that a considerable number of individuals willing to pay for internet upgrades reside in the southern parts of the United States, particularly in California, Florida, and Texas. The analysis also revealed that there is a significant difference between individuals who exhibit willingness to pay for internet upgrades and those who do not in terms of time spent watching sports on laptops.

The above findings have led to four main areas of recommendations: collaborations, target personas, business expansion, and seasonal offers. Collaborations with social network platforms, streaming music services, and mobile game developers, as well as video streaming companies, can attract a larger customer base to upgrade their internet services. Targeting young adults with full-time or part-time employment and families falling within the annual income range of 50,000 to 99,999 dollars will be a primary focus for sales. Expansion efforts should prioritize states with a higher willingness to pay, such as California, Florida, and Texas. Seasonal package offers, specifically targeting sports enthusiasts during popular events, can attract a wider customer base and enhance customer loyalty.

2. Introduction & Motivation

As a home internet service provider, HyperNet has been serving customers for several years. However, as technology continues to evolve and online activities become more complex,

customers are seeking faster internet speeds to meet their growing needs. To address this, HyperNet is considering offering an internet speed upgrade option to its customers.

As a team of data analysts working at HyperNet, we are committed to identifying the demand for the internet speed upgrade option in the market. Our analysis will address two key issues:

- understanding why customers want faster internet speeds, and
- identifying patterns in customer demand.

We will be analyzing Deloitte's Media Consumption dataset for this problem statement.
By providing insights into these issues, we aim to inform strategic decision-making for HyperNet and help the company remain competitive in the home internet service provider market.

This report aims to identify the factors driving customer demand for internet speed upgrades. By understanding customers' reasons for upgrading their internet speeds, HyperNet can tailor its marketing strategies and product offerings to better meet their needs. The report will analyze customer data to identify trends in demand for internet speed upgrades and explore potential marketing strategies to encourage customers to upgrade.

Ultimately, the insights provided in this report will help HyperNet make informed decisions that can lead to revenue growth, increased customer satisfaction, and loyalty. As a team of data analysts, we are committed to leveraging our skills and expertise to help HyperNet achieve its business goals and provide customers with the internet speeds they need to stay connected in an ever-changing digital landscape.

3. Data Description and EDA

We analyzed survey data from Deloitte regarding customers' media consumption habits for our endeavor. We chose the "data/DDS11 Data Extract with labels.csv" among 3 datasets of replies to a survey since it is the latest data and will reflect current trends better and assist in developing more accurate models. The dataset contains responses to questions about demographics, media owned or planned to be owned, media value ranked, time spent preference, media subscription, and entertainment habits. There are 196 columns and 2131 rows. The dataset's rows each correspond to a survey response, and each column either shows all possible answers to a single question or a binary choice among one question's options.

	record - Record number	Q1r1 - To begin, what is your age?	Q4 - What is your gender?	age - you are...	Q2 - In which state do you currently reside?	region - Region	QNEW3 - What is your employment status?	Q5 - Which category best describes your ethnicity?	QNEW1 - Do you have children living in your home (excluding yourself if you are under 18)?	QNEW2 - How old are the children in your home? -0-4 years	QNEW2 - How old are the children in your home? -5-9 years	QNEW2 - How old are the children in your home? -10-13 years	QNEW2 - How old are the children in your home? -14-18 years	QNEW2 - How old are the children in your home? -19-25 years	QNEW2 - How old are the children in your home? -26+ years	QNEW2 - How old are the children in your home?- Don't Know
0	4.0	36.0	Male	34-50	Georgia	South	Employed full-time or part-time	White or Caucasian (Non-Hispanic)	Yes	No	No	Yes	Yes	No	No	No
1	6.0	26.0	Female	20-26	New York	Northeast	Employed full-time or part-time	White or Caucasian (Non-Hispanic)	Yes	Yes	Yes	No	No	No	No	No
2	9.0	32.0	Female	27-33	New Jersey	Northeast	Employed full-time or part-time	White or Caucasian (Non-Hispanic)	Yes	Yes	No	No	No	No	No	No
3	11.0	25.0	Female	20-26	California	West	Employed full-time or part-time	White or Caucasian (Non-Hispanic)	Yes	Yes	Yes	No	No	No	No	No

Fig 1

2127	3591.0	70.0	Male	70 or older	Massachusetts	Northeast	Retired	White or Caucasian (Non-Hispanic)	No	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
2128	3620.0	18.0	Male	14-19	Alabama	South	Unemployed	White or Caucasian (Non-Hispanic)	No	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
2129	3610.0	79.0	Female	70 or older	Illinois	Midwest	Retired	White or Caucasian (Non-Hispanic)	No	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
2130	3673.0	77.0	Female	70 or older	Pennsylvania	Northeast	Retired	Multiracial	No	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!

2131 rows × 198 columns

Fig 2

There are five steps in this section, which are elaborated below.

3.1 Filter the Data and Define the Target Variable

We made the decision to use owning Home Internet access as a filter in light of the aforementioned issue. We eliminated rows whose responses to the question ‘Q26 - Which of the following subscriptions does your household purchase?-Home internet’ that were ‘No’. Home internet is used because people without access to the internet at home are not our intended market. We selected the column Q29: ‘Q29 - You said that you subscribe to home Internet access, how much more would you be willing to pay to receive double your download speed?’ as our target variable to determine who is willing to pay more for better Internet speed.

3.2 Perform a Demographic Analysis

We performed a demographic analysis on our target customers based on their age, gender, region, ethnicity, employment status, and annual household income.

3.2.1 Age & Gender

Most consumers are between the ages of 16 and 26 and 68 and 70. Compared to male customers, there are slightly more female customers.

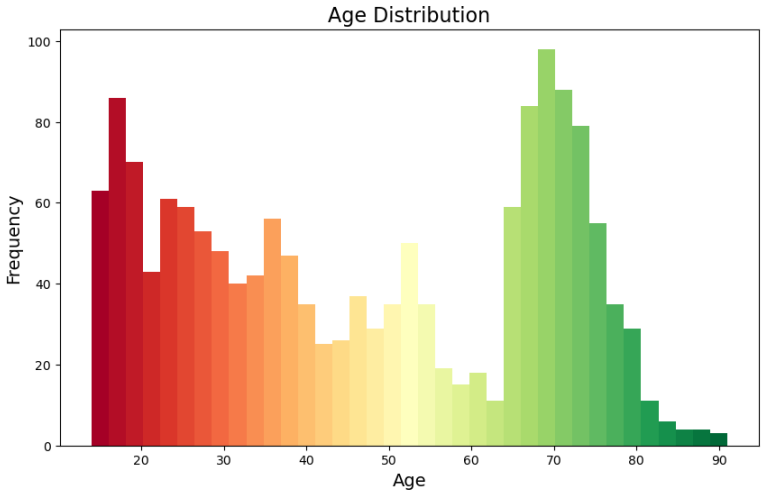

Fig 3

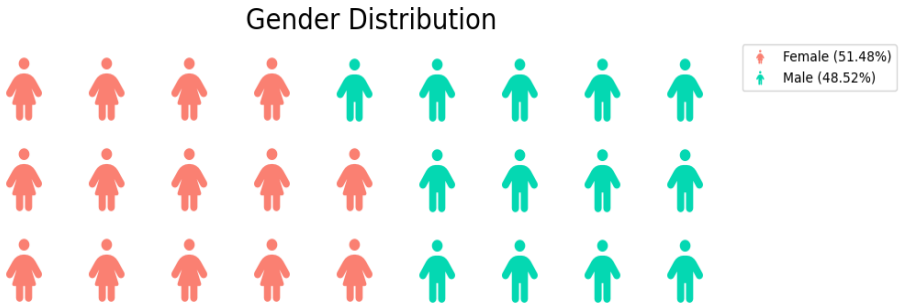

Fig 4

3.2.2 Region & Ethnicity

Most of the customers live in the South, while there is an even distribution of customers in the West, Northwest, and Midwest. Customers are predominantly White or Caucasian.

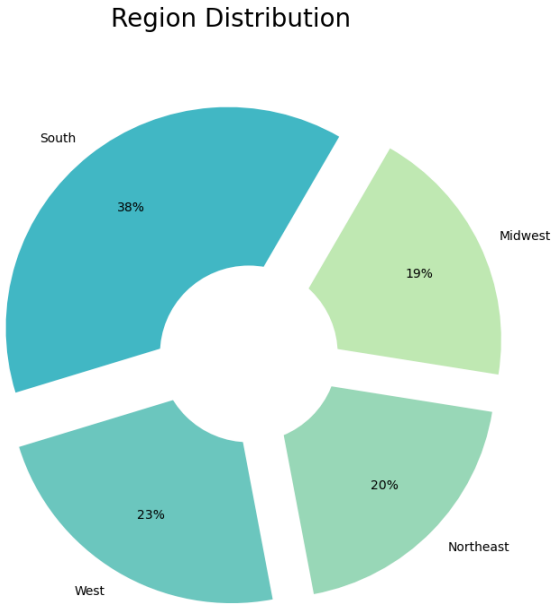

Fig 5

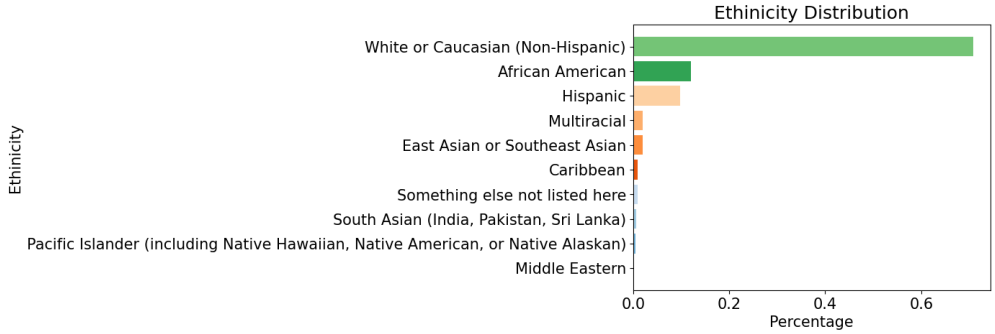

Fig 6

3.2.3 Employment Status & Annual Household Income

The majority of customers have full- or part-time jobs. The majority of the customers make between $50,000 and $100,000 per year in their households.

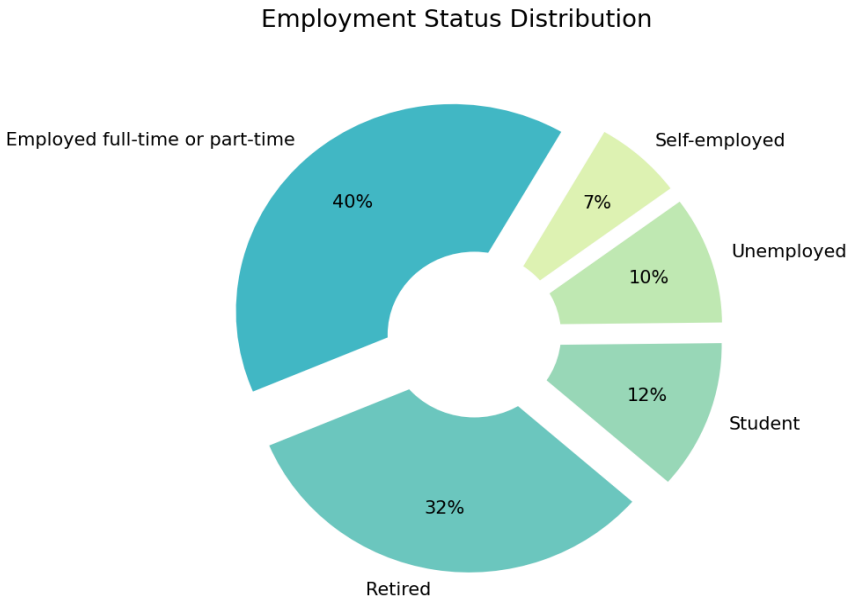

Fig 7

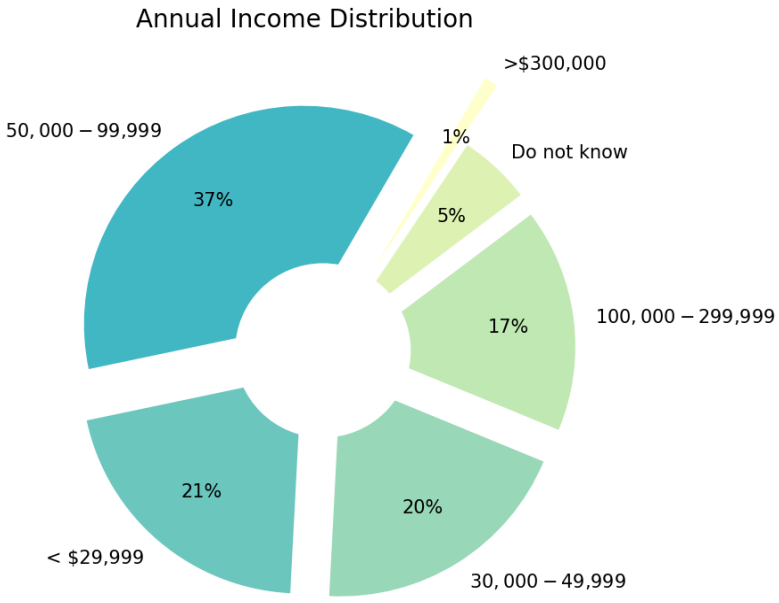

Fig 8

3.3 Handle the Missing Values

3.3.1 Drop Columns with More Than 60% Missing Values

	Missing count	Missing ratio
QNEW2 - How old are the children in your home?-0-4 years	978	0.627728
QNEW2 - How old are the children in your home?-5-9 years	978	0.627728
QNEW2 - How old are the children in your home?-10-13 years	978	0.627728
QNEW2 - How old are the children in your home?-14-18 years	978	0.627728
QNEW2 - How old are the children in your home?-19-25 years	978	0.627728
QNEW2 - How old are the children in your home?-26+ years	978	0.627728
QNEW2 - How old are the children in your home?-Don't Know	978	0.627728
Q11r1 - Flat panel television - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish to rank. Th	948	0.608472
Q11r2 - Digital video recorder (DVR) - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish to ra	1433	0.919769
Q11r3 - Streaming media box or over-the-top box - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selectio	1479	0.949294
Q11r4 - Portable streaming thumb drive/fob - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. M	1506	0.966624
Q11rNew1 - Over-the-air digital TV antenna (for free access to network broadcast without pay TV subscription) - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most value	1504	0.965340
Q11r5 - Blu-ray disc player/DVD player - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish to	1464	0.939666
Q11r6 - Gaming console - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish	1298	0.833119
Q11r7 - Portable video game player - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each	1514	0.971759
Q11r8 - Computer network/router in your home for wireless computer/laptop usage - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most va	1235	0.792683
Q11r9 - Desktop computer - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish to rank. The fir	995	0.638639
Q11r10 - Laptop computer - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the order you wish to rank. The fir	815	0.523107
Q11r12 - Tablet - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in the orde	1234	0.792041
Q11r14 - Dedicated e-book reader - Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued. Make your selections by clicking each item in	1507	0.967266

Fig 9

3.3.2 Drop Columns with Limited Information

The record number column and final weights column do not contribute to the whole analysis. We chose to drop these columns.

3.3.3 Fill in the Rest of the Missing Values Based on the Questions

In rating questions like ‘Of the products you indicated you own, which [totalcount] do you value the most? Please rank the top [totalcount], with "1" being the most valued’, we replaced missing values with 4 to show that users do not think this product is among their top 3 options.

We used Strings like "No Smart Phone" to fill in the blanks for binary questions like ‘What types of apps do you use frequently (everyday/weekly) on your smartphone?’ to substitute missing values to suggest that the population lacks the specific type of electronic device.

3.4 Apply Summary Statistics, Univariate & Bivariate Analysis

In order to run our summary statistics and analysis, we first separated our dataset into Numerical-Column-Only and Categorical-Column-Only datasets.

A portion of the Numerical-Column-Only dataset's summary is provided below. The code notebook contains the complete summary.

	Q1r1 - To begin, what is your age?	Q15r1 - Smartphone - Of the time you spend watching movies, what percentage of time do you watch on the following devices?	Q15r2 - Tablet - Of the time you spend watching movies, what percentage of time do you watch on the following devices?	Q15r3 - Laptop/Desktop - Of the time you spend watching movies, what percentage of time do you watch on the following devices?	Q15r4 - Television - Of the time you spend watching movies, what percentage of time do you watch on the following devices?	Q16r1 - Smartphone - Of the time you spend watching sports, what percentage of time do you watch on the following devices?
count	1558.000000	1558.000000	1558.000000	1558.000000	1558.000000	1558.000000
mean	47.308087	9.435815	7.626444	23.329910	47.284339	7.505135
std	21.493291	19.208739	17.758325	33.136552	43.848890	18.518152
min	14.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	27.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	47.000000	0.000000	0.000000	0.500000	50.000000	0.000000
75%	69.000000	10.000000	5.000000	40.000000	100.000000	0.000000
max	91.000000	100.000000	100.000000	100.000000	100.000000	100.000000

Fig 10

A portion of the Categorical-Column-Only dataset's summary is provided below. The code notebook contains the complete summary.

	Q4 - What is your gender?	age - you are...	Q2 - In which state do you currently reside?	region - Region	QNEW3 - What is your employment status?	Q5 - Which category best describes your ethnicity?	QNEW1 - Do you have children living in your home (excluding yourself if you are under 18)?	Q6 - Into which of the following categories does your total annual household income fall before taxes? Again, we promise to keep this, and all your answers, completely confidential.
count	1558	1558	1558	1558	1558	1558	1558	1558
unique	2	6	50	4	5	10	2	6
top	Female	70 or older	California	South	Employed full-time or part-time	White or Caucasian (Non-Hispanic)	No	$50,000 to $99,999
freq	802	369	168	593	618	1104	978	571

Fig 11

There are some outliers in the categories columns if we further examine. On every predictor, we do both univariate and bivariate analysis. Below are a few examples that illustrate the problems with outliers.

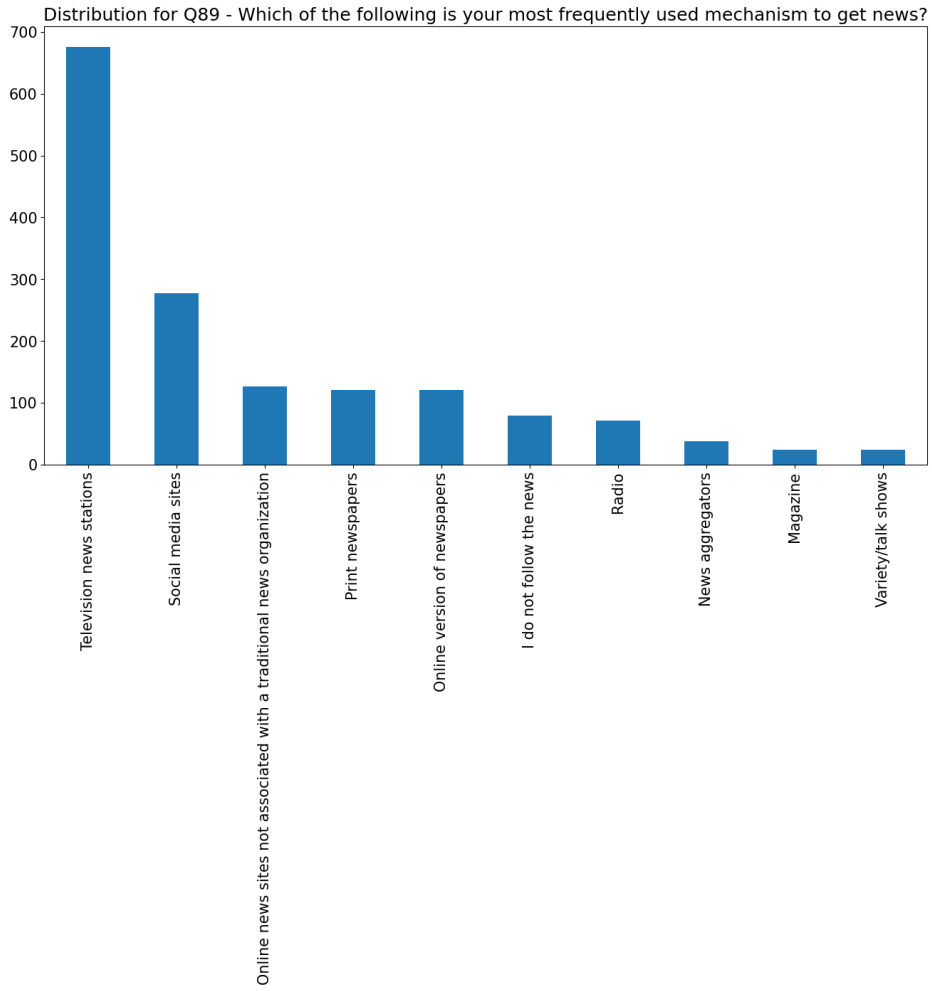

Fig 12

The above univariate analysis reveals that, in comparison to other mechanisms, few people use Magazine and Variety/Talk Show as their most frequent used mechanism to get news.

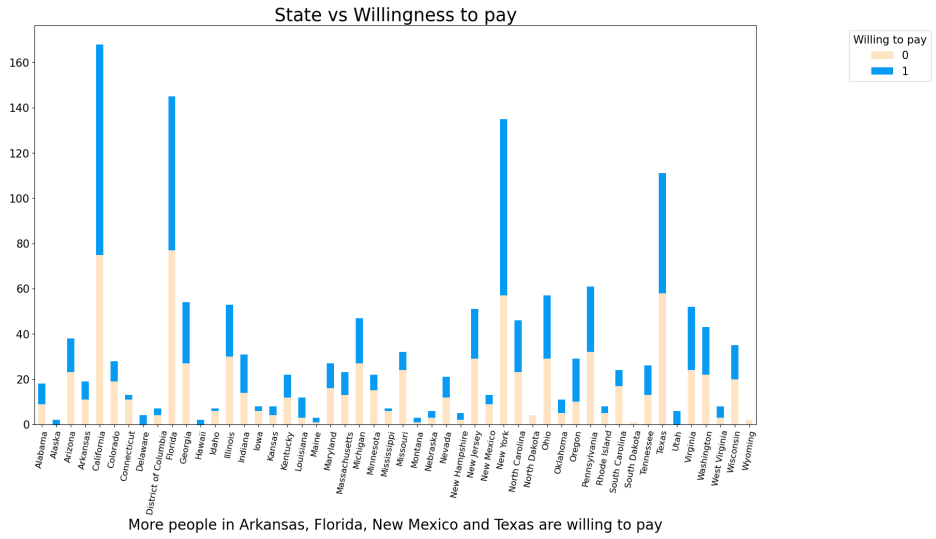

Fig 13

From this bivariate analysis, we can observe that practically everyone chooses "Willing to pay" for residents of states like Alaska, Delaware, and Hawaii. Because of this, those who select "Unwilling to pay" become outliers.

Even rarely chosen categories nonetheless contribute to the general distribution of the column data. We think that in this case, replacing or removing outliers could bias the data. As a result, we did not specifically address any outliers in this investigation.

3.5 Encode Categorical Data

First, we used binary encoding to convert every "Yes" to a "1" and every "No" to a "0".
In the following step, we changed the responses in our target variable column to "1" and "0," which signify the customers' willingness to pay (1) or lack thereof (0).
As a result, we used One Hot Encoding to transform categorical variables into a binary vector, where each member in the vector stands for a different category.

4. Model Development, Evaluation and Interpretation

In the Model Development and Evaluation phase, we tackled two problems: understanding the reasons behind customers' desire for faster internet speeds through Supervised Learning, and identifying patterns in customer demand through Unsupervised Learning.

4.1 Supervised Learning

For Supervised Learning, we used K-fold cross-validation to assess the performance of ten different machine learning models.

	Training Set	Validation Set
Naive Bayes	0.7439	0.7195
Logistic Regression	0.7042	0.6906
KNN	0.7865	0.7002
SVM	0.5738	0.5738
Decision Tree	1.0	0.6662
Bagging Decision Tree	0.9855	0.7163
Boosted Decision Tree	1.0	0.6752
Random Forest	1.0	0.7432
Voting Classification	0.9136	0.7233
Neural Network	0.7867	0.6630

Fig 14

Based on the training and validation accuracy, we found that most models were overfitting the data. Therefore, we decided to proceed with Logistic Regression for further analysis.

The Logistic Regression model was trained using Kfold on a dataset consisting of 1558 observations, with 744 belonging to the positive class and 814 to the negative class.

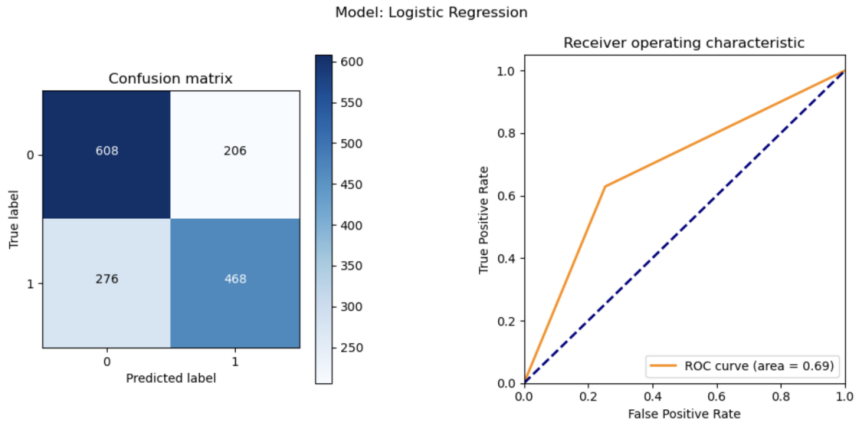

Fig 15

Analyzing evaluation metrics:

- An ROC area of 0.69 can be considered relatively good or acceptable since it represents a significant improvement over random guessing (ROC area of 0.5).

- A precision of 0.987 indicates that out of all the predicted positive cases, 98.7% were actually true positives, while only 1.3% were false positives. This suggests that the model has a high degree of confidence in its positive predictions, and is not making many false positive errors.

- A recall of 0.992 indicates that out of all the actual positive cases, 99.2% were correctly identified as positive by the model, while only 0.8% were incorrectly classified as negative. This suggests that the model has a high sensitivity to the positive class, and is not missing many true positive cases.

- Finally, an F1 score of 0.9899 indicates that the model has a good balance of precision and recall, and is performing well in terms of both correctly identifying true positives and avoiding false positives.

- Precision, recall and F1 score metrics suggest that the model is performing well in correctly identifying the positive class and avoiding false positives.
Next, we looked at Feature importance.

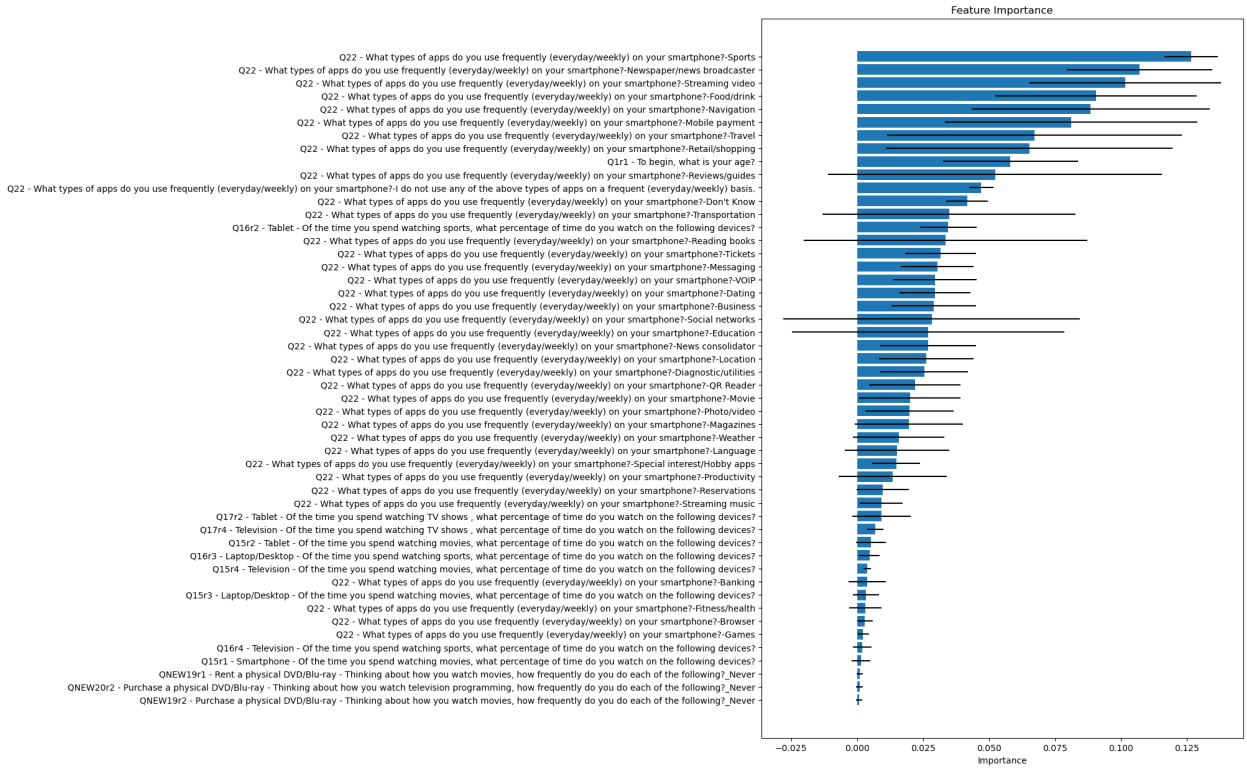

Fig 16

In our analysis, we evaluated the factors that influence customers' willingness to pay for higher internet speeds. We found that Q22, which relates to categories of frequently used apps, and age were the most significant factors in determining customer behavior in this context.

4.2 Unsupervised Learning

To gain further insights, we performed unsupervised learning techniques, specifically PCA with K-means and dendrogram with agglomerative clustering. We focused on the subset of customers who were willing to pay for higher internet speeds, allowing us to obtain a more targeted understanding of their characteristics and preferences.

We chose K-means for clustering because it is a simple and computationally efficient algorithm that produces clusters that are easy to interpret and visualize.

We implemented PCA to reduce the dimensionality of the data.

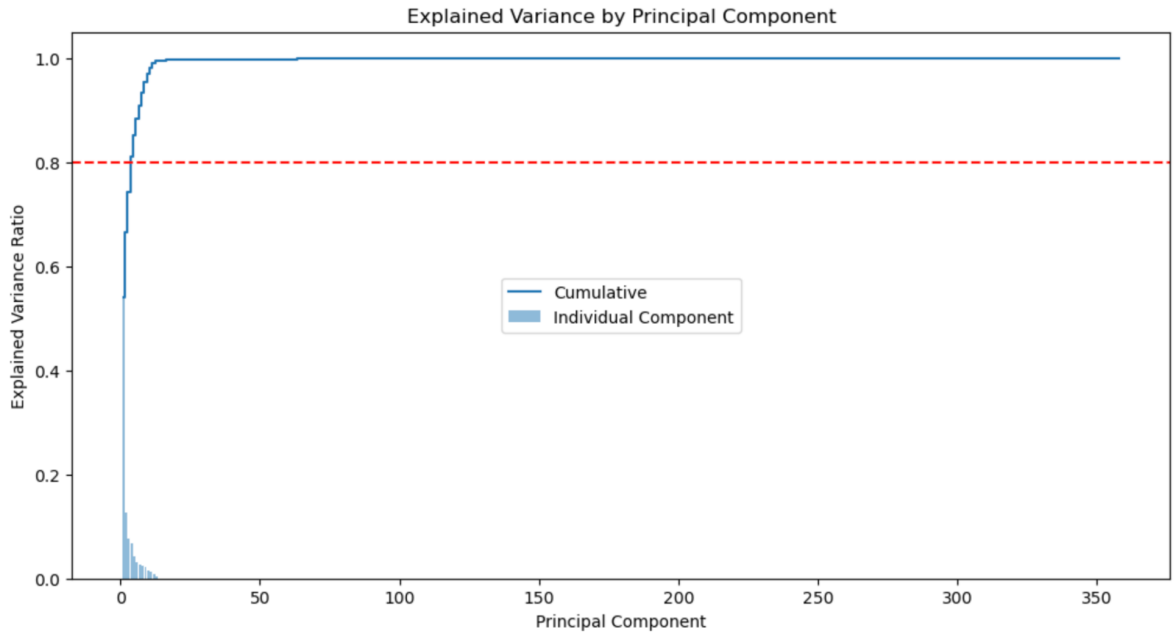

Fig 17

We selected 4 principal components since they can explain ~80% of the variance.

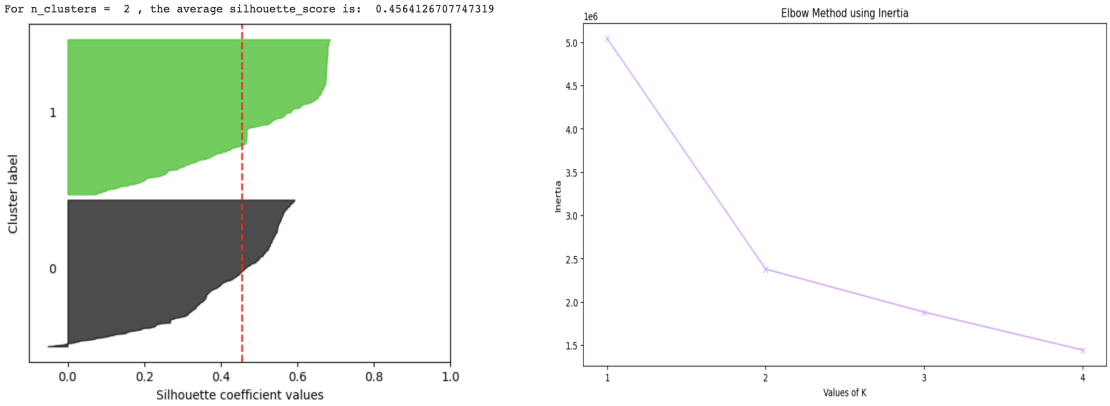

Fig 18

Using silhouette analysis and the elbow method, we found that the optimal number of clusters was 2.

On plotting PCA scores with respect to the clusters, we can see 2 distinct clusters.

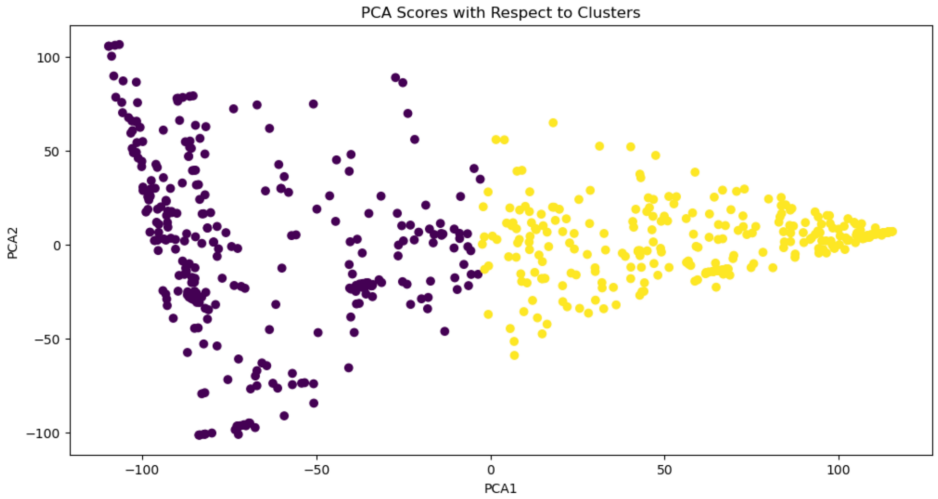

Fig 19

On plotting Age vs Cluster, we can see that we have mostly younger people in cluster 0 as opposed to cluster 1 which has a more distributed age range.

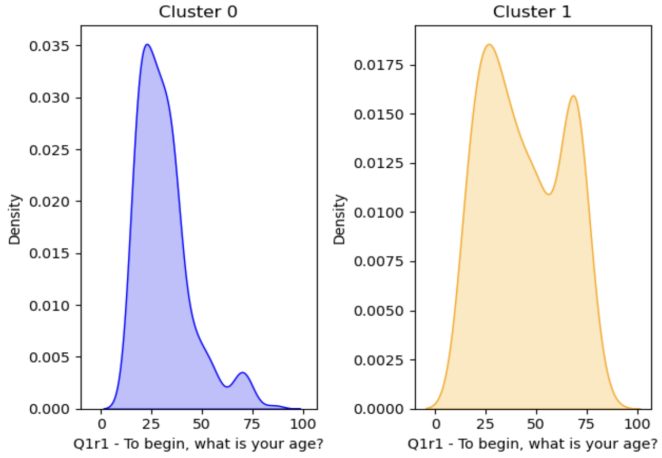

Fig 20

We also used agglomerative clustering, which is more flexible than K-means as it allows for different linkage methods and doesn't require specifying the number of clusters in advance.

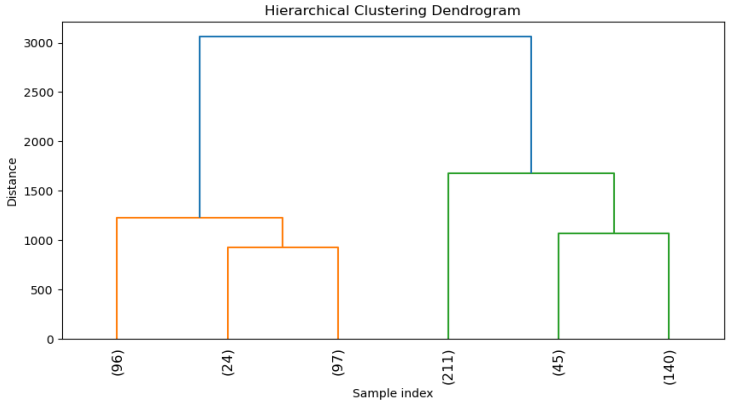

Fig 21

The dendrogram showed two main clusters with dissimilarity represented by the distance and height.

Similarly, on plotting Age vs Cluster, we can see that we have mostly younger people in cluster 0 as opposed to cluster 1 which has a more distributed age range.

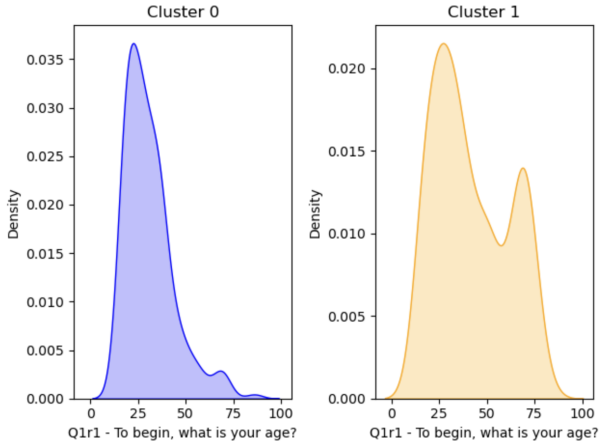

Fig 22

In both clustering methods, we used a random forest classifier to identify the distinguishing factors between clusters. We found that owning certain media or home entertainment equipment and age were the main factors that differentiated the clusters.

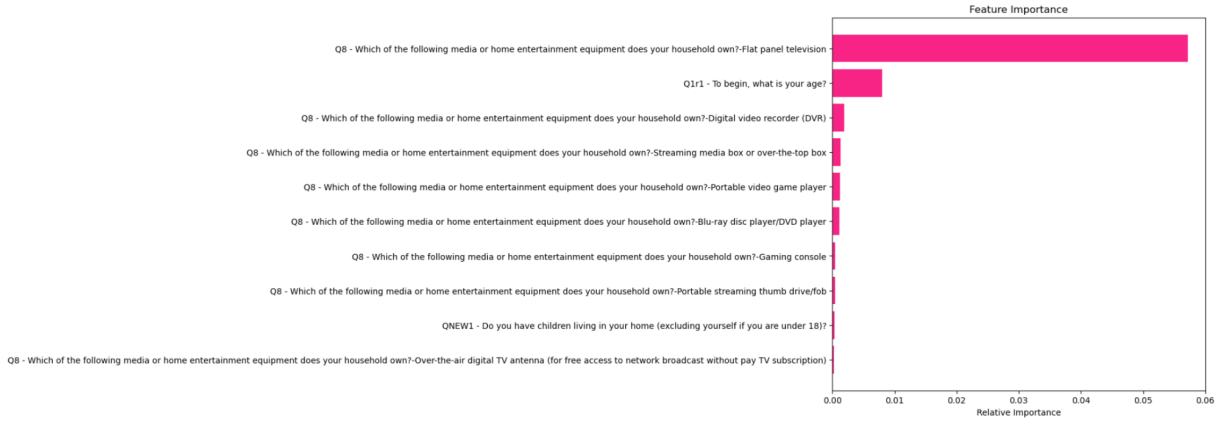

Fig 23

Overall, the consistency between K-means and agglomerative clustering and the similarity in distinguishing factors for each cluster indicate that the clusters are meaningful and robust. This segmentation of the customer base can inform marketing and business strategies tailored to the preferences and demographics of each cluster, ultimately improving customer engagement and satisfaction.

5. Recommendations

Based on our final model choice in the supervised learning part —— Logistic Regression, we figured out the features that are most strongly correlated with our target variable: whether or not a respondent is willing to pay more to double their home Internet access.

We conducted feature analysis by dividing the categorical or numerical variables into 2 groups: willing to pay and unwilling to pay on the top 8 most relevant features which are **App use on smart phones**, **Age & Employment**, **Region & States**, **Income**, **Time spent watching sports on Laptop**,**Time spent watching shows & movies on Tablet**.

4 main areas of recommendations are proposed based on the analysis on the above important features, including collaborations, target personas, business expansion and seasonal offers.

5.1 Collaborations

Upon examining the top 10 smartphone applications and the corresponding respondents grouped by willingness to pay, it is evident that individuals who are willing to pay exhibit higher usage frequency of social network platforms, streaming music services, and mobile games, compared to those who are unwilling to pay.

In light of these findings, potential collaborations may be pursued with companies operating in industries related to social networks, such as Instagram, Facebook, TikTok, and Snapchat, as well as streaming music services like Spotify and Apple Music, and smartphone game developers like Supercell, Niantic, and Innersloth. Such partnerships have the potential to attract a larger customer base to upgrade their internet services.

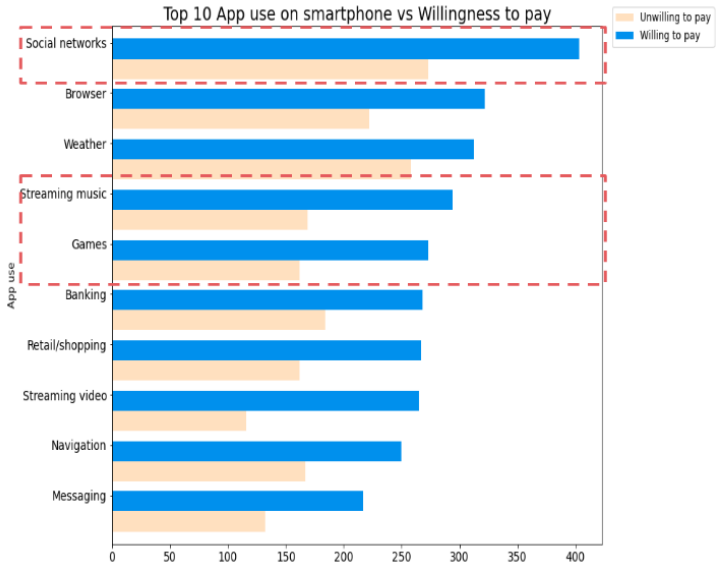

Fig 24

HyperNet's product team can implement customized Internet bundle services that are tailored to specific app usage on smartphones, where specific data packages in GB or MB will correspond to the intended app usage. As an example, a 10 GB package may be offered for listening to music on Spotify.

Analysis of the data gathered on tablet usage reveals a notable distinction between respondents who are willing to pay and those who are unwilling to pay. Specifically, individuals who are willing to pay tend to spend more time watching TV shows and movies on their tablets than those who are unwilling to pay.

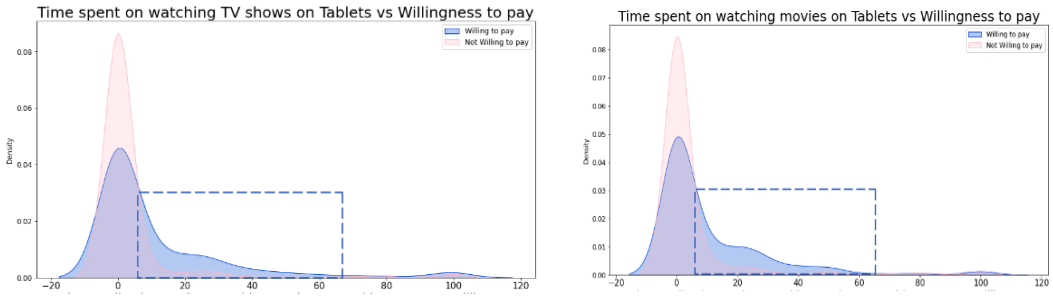

Fig 25

Building upon the aforementioned insight, collaborations with streaming video companies such as YouTube, Netflix, fuboTV, Disney+, and Amazon Prime Video may also be pursued. To achieve this, we propose a multi-pronged approach.

Firstly, customized internet bundle services similar to the ones described earlier will be offered, tailoring data packages in GB or MB to specific app usage. Additionally, we will sell internet packages bundled with video platform memberships, as an additional incentive to attract more potential customers. By offering this comprehensive service package, we aim to cater to the growing demand for quality internet and video streaming services.

5.2 Target Personas

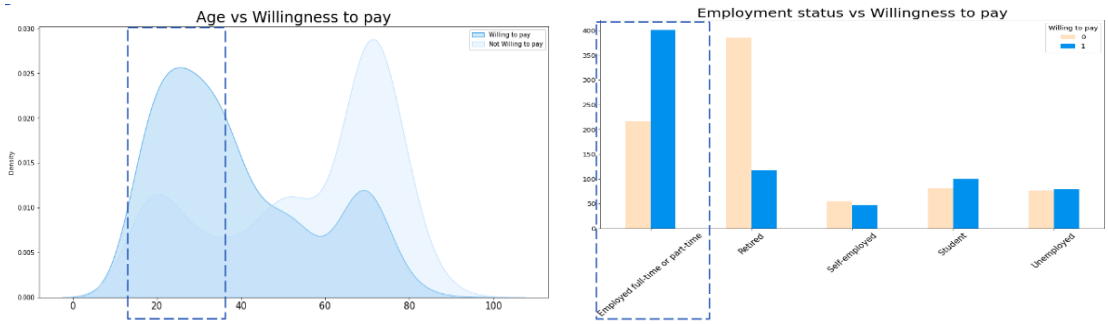

Fig 26

An examination of the personas of respondents who exhibit willingness to pay reveals that the majority of individuals fall within the age bracket of 16 to 36 years, with full-time or part-time employment.

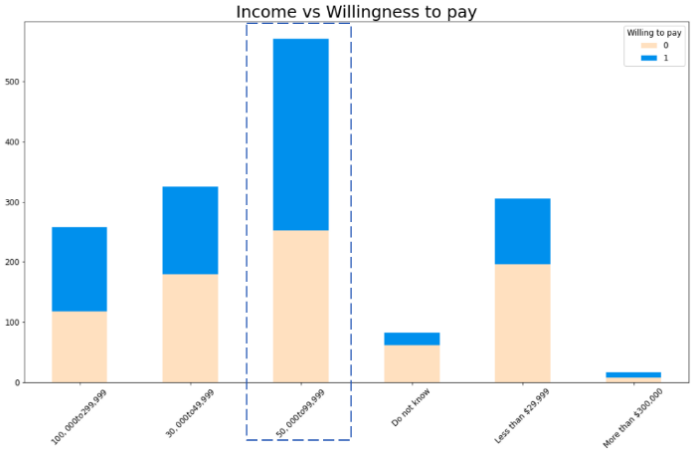

Fig 27

An analysis of the personas of respondents who exhibit willingness to pay suggests that the majority of individuals fall within the annual household income range of 50,000 to 99,999 dollars.

Considering the above findings, our sales team will focus primarily on targeting young adults who are predominantly employed full-time or part-time, through sales calls and emails. For household services, families falling within the annual income range of 50,000 to 99,999 dollars will be the primary target for service upgrades.

5.3 Business Expansion

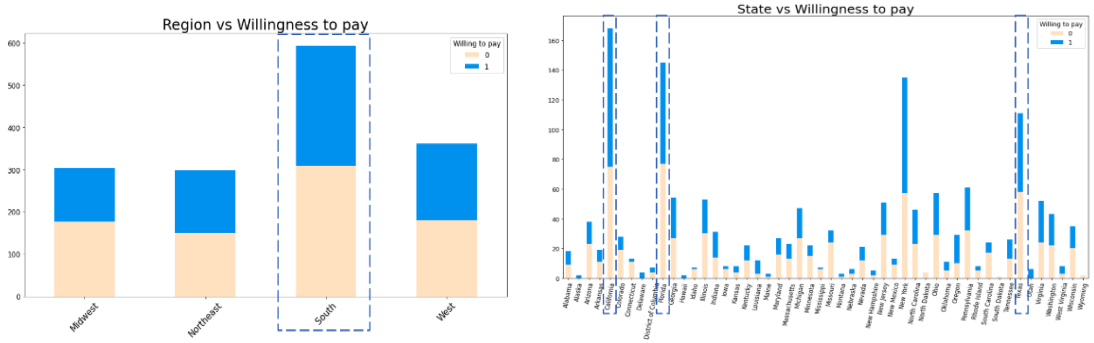

Fig 28

An analysis of the distribution of respondents who exhibit willingness to pay and those who do not, with a focus on the regional and state-wise distribution, has revealed that a considerable number of individuals who are willing to pay for internet upgrades reside in the southern parts of the United States, particularly in California, Florida, and Texas.

This insight is of great significance for HyperNet's expansion plans, as it highlights the potential for targeted marketing efforts in these states. The observed differences in demand suggest that expanding into these regions would be a high priority for the business. As such, the expansion strategy may be tailored to cater to the specific demands of customers in these states and maximize the potential for revenue growth. Additionally, identifying the unique requirements of the target market in these regions and customizing service offerings accordingly may provide a competitive advantage for HyperNet.

5.4 Seasonal Offers

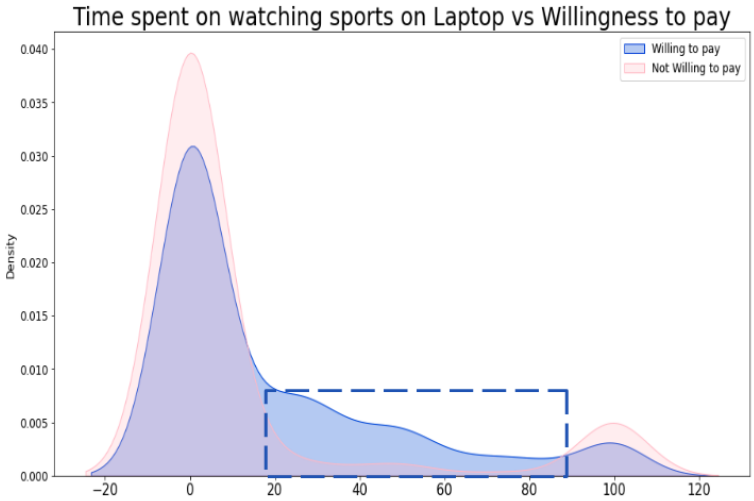

Fig 29

Upon analyzing the time spent by respondents on watching sports on a laptop, it is apparent that there is a significant difference between individuals who exhibit willingness to pay for internet upgrades and those who do not. Specifically, respondents who are willing to pay tend to spend more time watching sports on laptops than their unwilling counterparts.

In light of this finding, HyperNet's product team can consider developing new package choices for seasonal internet surfing service offers. To cater to the needs of sports enthusiasts, we could introduce special package offers during the season of highly popular sports events such as the Super Bowl, NBA Finals, and Iron Man. By aligning our services with the preferences of sports enthusiasts, we could attract a wider customer base and potentially increase revenue.

Furthermore, the development of targeted packages and promotional offers could also be an effective strategy for customer retention. By providing tailored services that align with the preferences of specific customer groups, we could establish ourselves as a preferred service provider, thereby enhancing customer loyalty and satisfaction.

6. Conclusion

In conclusion, our analysis using Logistic Regression showed that the frequency of app usage and age were the most significant factors in determining customer behavior when it comes to paying for higher internet speeds. This insight can help internet service providers better understand customer needs and tailor their marketing and business strategies accordingly.

Furthermore, our Clustering analysis using K-means and Agglomerative clustering revealed meaningful and robust customer segments based on their internet usage behavior. By leveraging these segments, businesses can develop targeted strategies and services to improve customer engagement and satisfaction.

The logistic regression model analysis identified key features strongly correlated with willingness to pay for internet upgrades, including app usage on smartphones, age and employment status,

region and state, income, and time spent watching shows and movies on tablets. Based on these findings, the team proposed four main recommendations: collaborate with companies in related industries, offer customized internet bundles based on app usage, target young adults employed full-time or part-time, and expand into the southern parts of the United States while developing seasonal offers. By implementing these recommendations, HyperNet can increase revenue growth, retain current customers, and attract a larger customer base. Moreover, customizing services according to target market requirements may provide a competitive advantage for the company.

Overall, our analysis demonstrates the importance of using both supervised and unsupervised learning techniques to gain valuable insights into customer behavior and preferences.

7. Work Distribution for the Project

Here is a work distribution summary for the Deloitte Survey Analysis project for HyperNet:

Ruchi: Model training, testing, and explaining (mainly responsible)
Caroline and Maxine: EDA and data preparation (mainly responsible)
All three team members: Collaborated to figure out recommendations based on model and EDA results.

It is worth noting that everyone participated actively during the entire project and collaborated effectively with each other to deliver high-quality work. Overall, it seems like a well-balanced distribution of work, with each team member contributing their skills and expertise to the project's success.