

DS Take Home Assignment

- Komal Mistry

Dataset

The dataset provided is a random and de - identified sample of mobility subscribers.

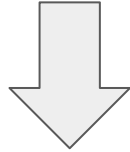
In general it has the following information about a customer:

- Customer IDs
- Customer Location Data
- Customer Demographic Data
- Mobility Subscription Info
- Tenure and Billing information
- Flag : Internet and TV

Task Description

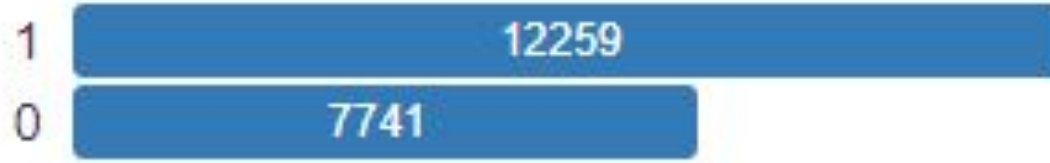
Build a model (or models) and provide corresponding recommendation(s) to help the marketing team launch the most effective campaigns.

- There are existing campaigns and offers for internet that can be leveraged or new campaigns can be launched
- There is a **limited campaign budget** (i.e. not all recommendations can be implemented) and **most campaigns have a capacity limit** (i.e. only select subscribers can be contacted)
- The **target audience** is **existing subscribers who do not have internet currently**



- Model needs to be used to learn more about mobility customers who don't have internet yet.
- Task at hand is to create a list of targeted customers who should be marketed to.

Problem Definition



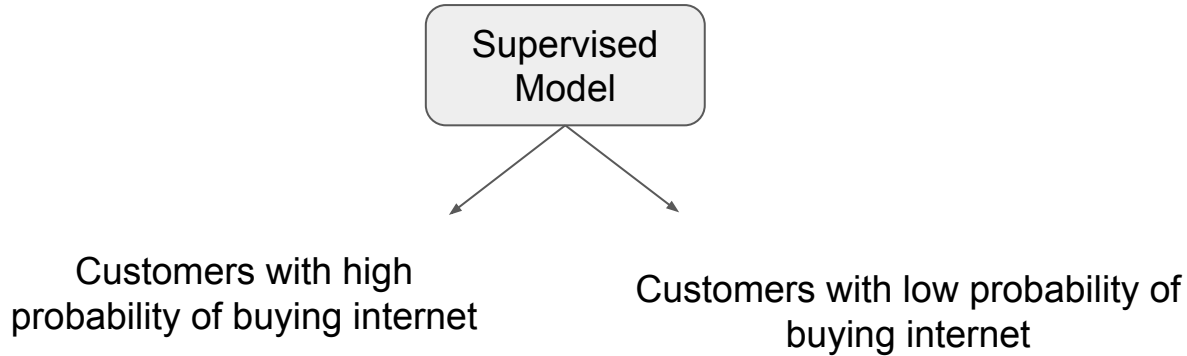
Above shows that 7741 customers in our dataset don't currently have internet.

Therefore, we want to convert this segment of people into internet customers.

No constraints : Advertise most popular internet plan to all 7741 customers.

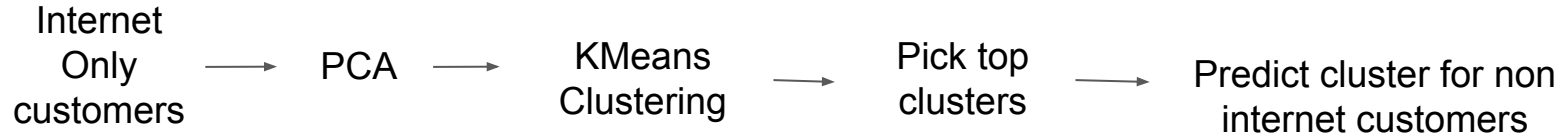
Constraints : Create targeted list of customers most likely to buy and suggest the most suitable internet plan for this smaller set of customers.

Process - Step 1



- This will help filter out customers with very low $P(\text{internet})$ and will help reduce the number of customers to be marketed to.
- But this doesn't provide any info on what should be marketed to each customer.

Process Step - 2



Once clusters are predicted for each customer, we can calculate the mode (i.e most frequent internet plan) in that cluster and recommend the same to the non-internet customer under consideration.

Therefore this approach helps us do the following:

- Customer segmentation wrt mobility info
- Find out which internet plan should be advertised to the non-internet customer based on what other similar customers liked.

(Assumption : We can find info about internet plan bought by existing customers)

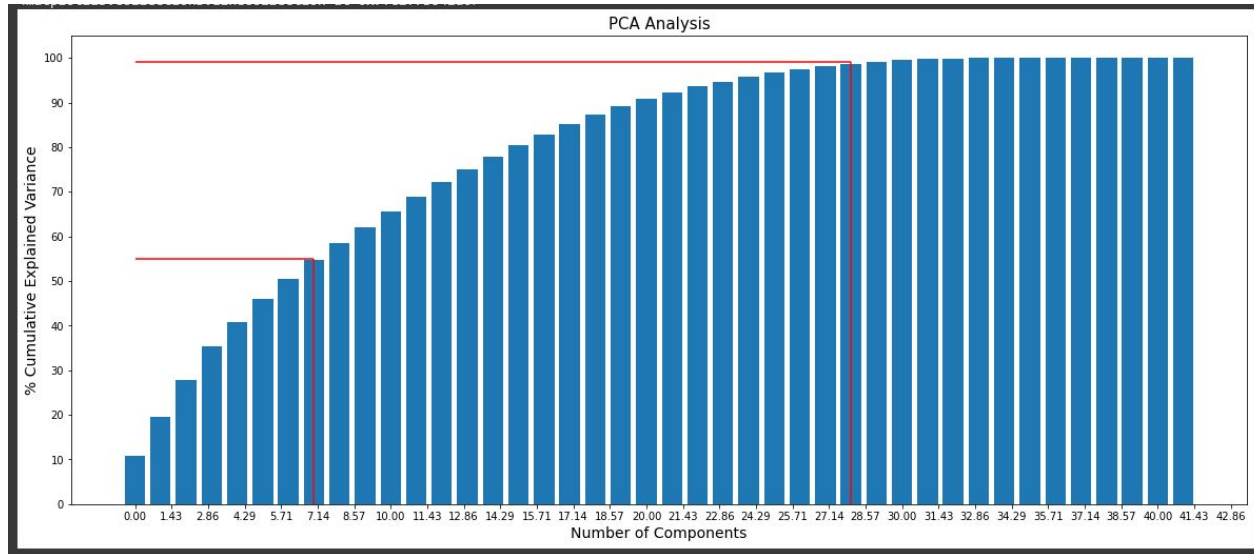
Step 1 : Supervised Classifier

Types of Models Considered:

1. `LogisticRegression`
2. `RandomForestClassifier`
3. `AdaBoostClassifier`
4. `GradientBoostingClassifier`

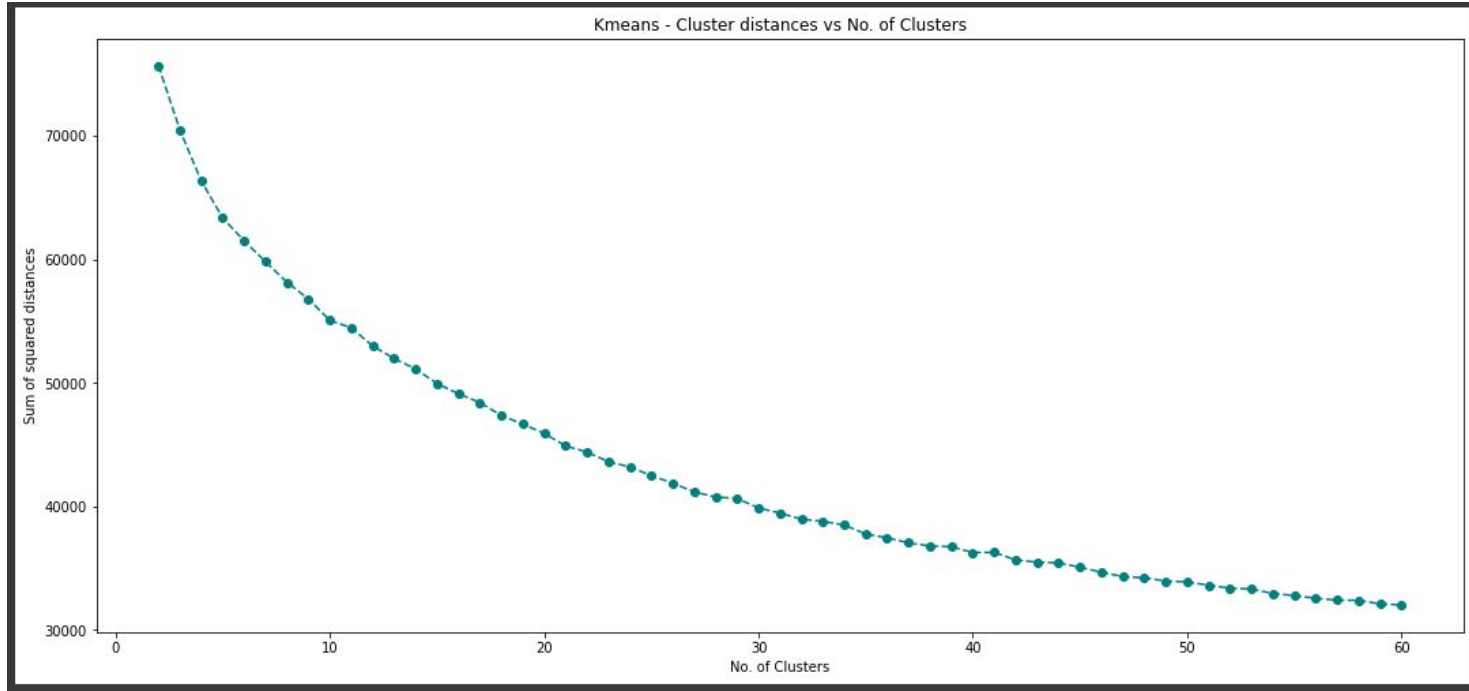
Step 2 : Clustering

Why did we pick 28 dimensions?



Step 2 : Clustering

Why did we pick 25 clusters ?



Suggested Plan

Using 80/20 approach / Pareto Principle, we can identify the targeted list of customers as follows:

1. Using supervised classifier divide non-internet customers into 2 groups using threshold at 0.8 to get only the most confident samples.
2. From clustering pick top k clusters which correspond to 20% of all internet customers [approx 2500 customers] -> pick the largest clusters

Picking the largest clusters -> targeting the largest customer segment first

3. Using non-internet customers from step 1, predict clusters for each.
4. Only retain non-internet customers who got assigned to the top - k clusters.
5. Determine the most popular internet plan in each cluster, recommend the same to the non-internet customers who are assigned to that cluster.

A/B Testing

Control Group

Contact random set of 1400 customers with most popular internet plan

Test Group

Contact identified target set of 1400 customer with their cluster assigned internet plan recommendation

Set up

1. Each customer gets advertised by email, with a link to buy the internet plan recommended.
2. We can track the click through rates for all sent emails, to determine which customers responded +vely to advertising.
3. Control and Test groups can work individually for a week, after which we can compare click through rates for each group to assess effectiveness of the DS enabled campaign mgmt.

Future Plan

- Once A/B Testing validates the effectiveness of DS approach, we can start easing out the 80/20 approach.
- First we can include other smaller clusters into top-k clusters, this will help release advertising to the smaller customer segments as well.
- To further increase number of customers to be advertised to, we can lower the threshold of supervised classifier to include customers who rank lower in $P(\text{Buying internet})$.

Extra (Using mobility plan info)

- When determining the internet plan suggestion for non-internet customer, we can make use of the PRICE_PLAN feature.
- Using PRICE_PLAN we can extract information about mobility plan details i.e number of GB provided.

0 GB
2-5 GB
5-10 GB
10-20 GB
20 + GB

Non- internet
customer

5GB mobility plan

0 GB
2-5 GB
5-10 GB
10-20 GB
20 + GB