INTER HALL DATA ANALYTICS

TEAM 11

Index

1 INTRODUCTION

An interbank network, also known as an ATM consortium or ATM network, is a computer network that enables ATM cards issued by a financial institution that is a member of the network to be used to perform ATM transactions through ATMs that belong to another member of the network. Interbank ATMs are particularly useful in cases the ATM branch of the same bank is not nearby and more importantly they allow customers to have surcharge (a fee deducted by the banks on transactions from non-bank ATM cards) free transactions in the same network.

MoneyPass, AllPoint, Co-Op are three important ATM networks in the United States widely distributed across the country. They have various smaller group of banks and credit unions under their name. American Chartered Bank, U.S. Bank under MoneyPass, Capital One 360, Whitney Bank under AllPoint, 1st Advantage Federal Credit Union, Workers Credit Union under Co-Op are examples of some banks. These ATM networks have also collaborated with bigger organizations like 7Eleven, WalMart etc. to install ATMs at their stores to enhance customers convenience. Hence, comparing their competitive strength is a motivating task in order to analyse who has a strategical edge over others in the market.

> At large, the study of revenue generation of ATMs poses an interesting yet complex research problem, and analysing it would help to unearth a myriad of contributing factors.

The ATM networks always try to maximize their return on investment by planning strategically their ATM network so as to maximize the number of transactions. The banks usually keep into account the following factors like Estimated traffic per day, Location popularity, Weekend or Weekday traffic, Rough estimation of transaction size in such a location, Service Routes (for Security company, is the ATM within a route or away from it), Other ATMs in the vicinity, Nearest Bank ATM Leasing space issues / costs, Marketing potential, Communication & Power infrastructure, Threat scale for vandalism and/or crime, Ease of servicing an ATM Brand Management (Exposure), Bank Owned ATM or Leased Through 3rd Party, Competition Factor (if there are multiple ATMs in the same location) etc. Using these factors the network tries to place an ATM. The companies also consider the cost of setting up an ATM in such locations and the amount of return that can be estimated. At last the companies can analyse whether the entire venture is successful.

The problem statement defines a similar problem where the ATM locations of three competing locations are mentioned in the region of California. The location features like the population density, average income, living standard can be gathered from the given zip-codes. The task requires us to form a logical inference of the features at hand in order to assign accurate priority to the appropriate features. After the assignment, we can sort regions in California where maximum transactions are possible then we can select the ATM network which has a larger number of branches in high priority regions.

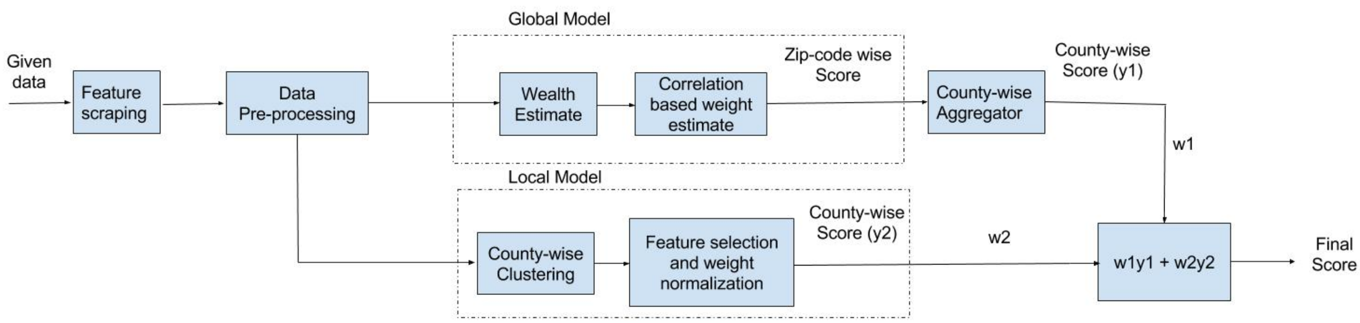
2 METHODOLOGY

Our entire task is divided into the following sections: data pre-processing, visualization of extracted features, inferring the priority weights to be assigned to each features and deduction. Post this, we exploit the weights to fit a regression to compute revenues generated by each ATM.

2.1 APPROACH

The estimation process for the revenue generation for each bank network involves considering two models. The function of each model is to give weights to features depending on the extent it affects the prediction. The first model named as the global model is used to find out the feature vector which influences the entire dataset. The local model is used to estimate the feature vector which has more influence on the prediction in the given county only. We use the "county-wise score" from both the estimators and weigh the outputs. The local model outputs are given more preference over the global model. The global model output is taken into account in order to avoid over-fitting in a particular county.

Figure 1: Estimation Model



The block diagram shown above in the Figure 1 describes the work-flow of the estimator which starts from the given data and feature extraction. The entire feature set is fed to both the estimators. The first assumes a function based on a minimal set of features and relates the weights of the other features based on the correlation to the function. The second estimator uses clustering among data present in a particular county. It takes by default seven clusters and assigns weights according to their influence on the hypothesis function.

2.2 DATA PREPROCESSING

1. *Feature Extraction*

   We have used the result zip for ATM available and scraped data relating to the zip-code from the website United States Zip-Code. The raw features available on the website per zip-code were scraped. Few of the important features are Population density, Housing units, Median home value and Families vs. Singles.

The total number of raw features extracted from the United States Zip-Code website is 155. The raw extracted data is pruned in the following data processing steps for better visualization purposes

2. *Latitute & Longitude Extraction*
   The latitude and longitude of individual zip-code is crawled using a Python Library, `Pygeocoder`, which exposes a convenient API for the above task.

3. *Frequency of ATMs*: The number of ATMs in a particular zip-code is calculated. This depicts the ATM density in a particular region. Various other inferences can be drawn regarding the advantage a particular ATM has over others depending on the population density in the area, and the overall status of people residing in the area.

4. *Nearest Zip-code computation*
   Data relating to the nearest zip-code sorted by distance was also collected from the United States Zip-Code website. The nearest zip-code provides insight into the likelihood of people using an ATM in a nearby region.

5. *Labelling the ATM name tag*
   The ATM name tag representing the city, shopping mart or gas station in which the particular ATM is located. The ATM name tags play a pivotal role in deciding the number of transactions as it provides an overview of the number of transactions. The ATMs present in shopping marts or gas stations are likely to have a higher number of transactions. From the label tag about each ATM we came up with a scoring scheme for each ATM.

   Name-tag classification of ATMs

Classification	Relative Score
Shopping Malls	10
Banks/ Exchange Centre	9
Recreation Centre	8
Gas Stations/Car wash	7
Office Area	6
Individual Store	5
Null Data	4

6. *Average Score per zip-code*: We gathered data for each unique zip-code then found out the average score per bank based on the ratings of the name-tag. The average score is further used to get the score of a bank network in a given county. This score is the basic building block in forming a conclusion regarding the strategic advantage of a bank network.

3 VISUALISATION

In this section, we attempt to unearth the relation between various features and revenue generation by the ATMs for each county.

In this particular example, to account for simplicity of our model, we assume that the **cost per transaction is uniform** across all the ATM networks.

Using this, we can reduce the problem to finding the **total number of transactions** for each county.

FRAME 1: WEALTH ESTIMATE

Since the data pertaining to number of transactions is not explicitly provided to us, we proceed by computing another measure, called the Wealth Estimate (WE) for each zip code, which is related as follows:

$$WE = PD \times MHI \times (1 - \%not\ earning) \tag{1}$$

Where,

1. *PD*: refers to the population density of the zip code
2. *MHI*: refers to the median household income of the people residing in the zip code
3. *% not earning*: refers to the percentage of people not earning in the zip code

Intuitively, we posit that the WE score could be a strong measure of the total number of transactions done by an ATM in that Zip Code. The visualizations rendered henceforth utilize this assumption and attempt to find relations between several demographic features with this metric.

Thus, the next section deals with the association of some relevant features with the Wealth Estimate, and the inferences we can draw from them. Some key points to note from the data are:

1. Out of all the counties, San Francisco and Los Angeles have the highest wealth estimates.
2. Out of all the networks, it is Allpoint which has the maximum number of ATMs in almost all counties.
3. Out of all the counties, Los Angeles has the highest number of ATMs.

3.1 ILLUSTRATIONS

To start with, we can turn our focus into seeing the frequency of ATMs in the 3 networks, i.e. Allpoint, Moneypass and Co-op across each county in the following figure **??**. Let us look at some of the features which were correlated with the Wealth Estimate score, and hence can be considered as relevant features.

1. *Transport*:

   Figure 2 demonstrates the correlation between the Wealth Estimate, and the cumulative percentages of people who commute via public transportation modes, and using cabs and motorcycles.

   (a) It was observed to have a correlation score of 0.6088, and it should be noted that this population consists of individuals older than 16 years old.

   (b) Further, since the WE is our estimate of the revenue generated by an ATM in the area, it can be inferred that as the percentage of people using public transport increases, the traffic increases, and thus the utilization of ATMs placed in such locations is greater and hence positively correlated.

2. *Rented 1-Bedroom Houses*: Figure 2 demonstrates the correlation between the Wealth Estimate, and the percentage of occupied houses which have only one bedroom.

   (a) The pair was calculated to have a correlation score of 0.467, which presents a positive correlation between the Wealth Estimate and the above variable.

   (b) Additionally, by this relation, it can be deduced that the revenue generation of an ATM increases as the percentage of people occupying single bedroom houses increases, due to the increase in the number of transactions per housing unit.

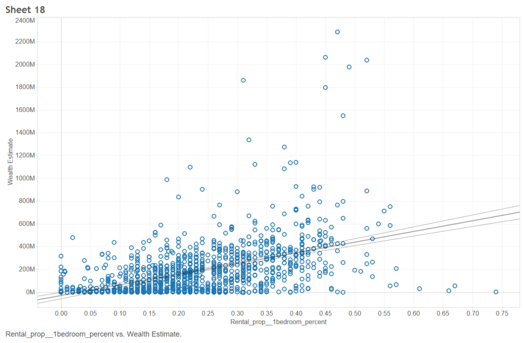
3. *Median Home Value*: Figure 2 demonstrates the correlation between the Wealth Estimate and median home value of the area represented by the zip code.

   (a) The pair was calculated to have a correlation score of 0.45, which presents a positive correlation between the Wealth Estimate and the above variable.

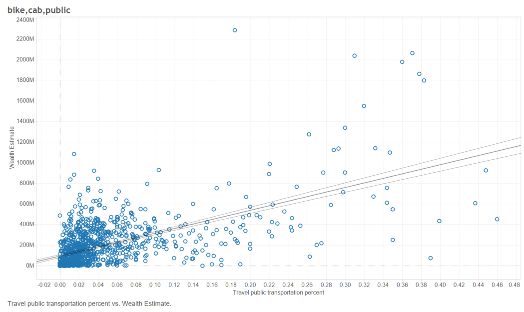
   (b) The median home value depicts the median of the value at which the homes are priced in the area. The price in a way reflects the overall economic status of the local area, and thus it would be more prudent to place an ATM in such an area which would guarantee copious transactions.

4. *Employed Sections*: Figure 2 demonstrates the correlation between the Wealth Estimate and the percentage of people employed in a full time or part time job.

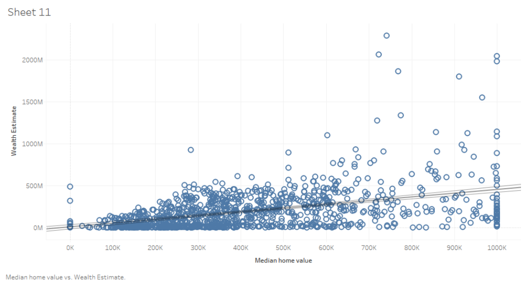
   (a) The pair was calculated to have a correlation score of 0.41, which presents a positive correlation between the Wealth Estimate and the above variable.



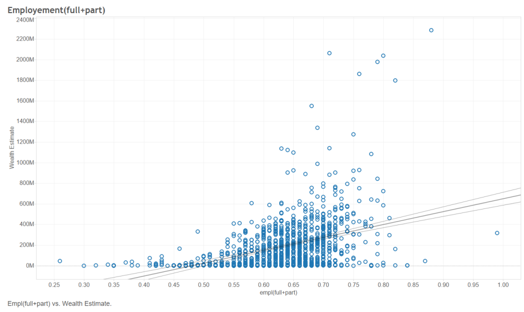
(a) Wealth Vs Rent For Single Bedroom



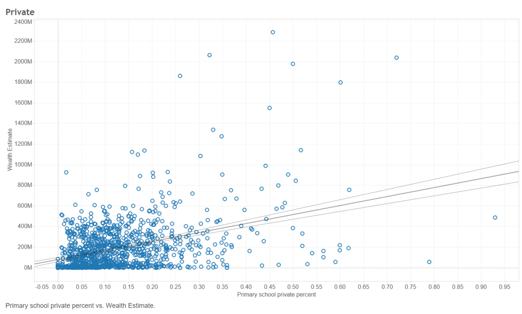
(b) Wealth Vs Transportation by public vehicles



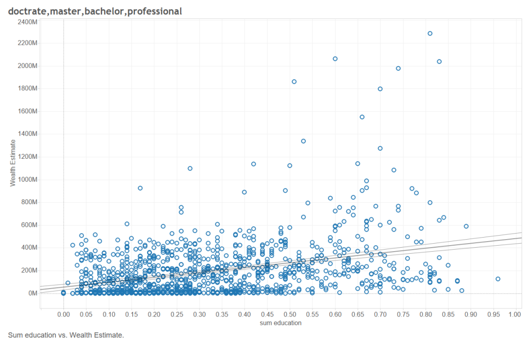
(c) Wealth Vs Median Home Value



(d) Wealth Vs Full & Part time Employment



(e) Wealth Vs Students enrolled in Private School



(f) Wealth Vs Persons with degree $\geq$ Bachelors

Figure 2: Plots of Wealth Estimate Vs Different Features Used in the Model

(b) As the percentage of employed people in the area goes up, the number of transactions is bound to increase to account for their expenditures in basic living amenities and sustaining a quality of life.

5. *Private Primary Schools*: Those areas which had a greater percentage of students in the age bracket of 3-17 years enrolled in private primary schools were observed to have higher wealth estimates, again alluding and augmenting the overall economic status of the area. We can draw a few inferences from this observation.

   (a) The education standard of the children are a reflection of the overall education status of the society. It can be related that families having children studying in private schools are more affluent thereby correlating to a higher wealth estimate. This is in turn corresponds to the number of transactions.

   (b) The education level provided in private schools can be considered to be better with respect to the fact that their students are much more stable financially.

6. *Educated Sections*: Places which had a higher percentage of educated people, including people who had Bachelors, Masters, Doctorates and other professional degrees were observed to have higher wealth estimates. This also turns out to be an obvious corroboration.

4 MODEL

In order to analyse accurately the variation of the number of ATM transactions we will consider two different approaches. One of them will be the global model where the decided weights are assigned according to the zip-code data over the entire dataset. This model gives us an idea about the factors which affects the wealth estimate globally. The local model is based on the factors which affect the wealth estimate in a particular county only. In general it is expected that every county will have similar features with high correlation as in the global model with some exceptional features which affect the prediction in that particular region only.

4.1 GLOBAL MODEL

The task of forming a proper global model involved manual inspection of the data features at hand and intuitively estimate whether it can remotely affect the number of transactions. For the given problem as we do not have a definite function we started off with an initial function of our own consisting of only the most important features (Reasoning for the features are to be provided later). Our initial model as stated in the visualization section is as follows:

$$WE = PD \times MHI \times (1 - \%not\ earning) \tag{2}$$

where,

1. *PD*: refers to the population density of the zip code
2. *MHI*: refers to the median household income of the people residing in the zip code
3. *% not earning*: refers to the percentage of people not earning in the zip code

From the visualization section we calculated the correlation between different features.

Taking into consideration only factors having a significant positive correlation with the wealth estimate, the features are given weights according to its correlation value with the wealth estimate.

The weights are normalized in order to calculate the score of each zip-code using Equation 3. The weight distribution for the significant features are mentioned in the table below.

Correlation based weight distribution

Feature	Correlation	Normalized weight
% Singles in Population	0.29	0.07
% of Singles with roommate	0.34	0.08
Travel Cab + Motorcyle + Public	0.61	0.14
Full Time & Part time employed	0.41	0.09
% in Private School (3-17 yrs)	0.39	0.09
Median Home Value	0.45	0.10
% Rental one bedroom	0.47	0.11
% Educated Section	0.35	0.0802
Population Density	0.79	0.18
Median House Income	0.21	0.048
$(1 -$ % not earning)	0.41	0.09

After computing these weights, we extract only the above corresponding features X, and apply a linear combination using the above weights w and Equation 5.

$$w_{i,norm} = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

$$y_{global} = \mathbf{w_{norm}}^T X \tag{4}$$

The relevance of the above features have been discussed in the previous section. The score y_{global} is combined with y_{local} to produce the final score. This will be discussed in our next section.

4.2 LOCAL MODEL

In order to shed more light on the local features of every zip code, we designed the local model. There would be some features that govern ATM revenue generation on a county level, and these may be more abstract and high level than the granular features that the zip code level analysis has to offer.

1. For analysing this, we partitioned the data by county, thus generating 55 data-sections. For each of the data sections, we applied a separate k-means clustering algorithm to group similarly behaving features closer to each other in high dimensional space.

2. After some experimentation and trade-off between computation and relevance, we settled on a $k = 7$. Consequently, this resulted in allotting 7 labels y to each of the zip code based samples for each county.

3. To reverse engineer the feature importance, the output labels, y, thus generated are used as a class label for supervised classification using a random forest classifier.

4. This enabled us to reveal the relevance of each feature and hence interpret the feature importance as the weight w_i. The weights could further be normalised and exploited as exactly as in 3 and 5 respectively.

The k-means clustering method is an iterative clustering algorithm, relies on the convergence of data points to group themselves in k clusters, which are decided by the nature of the data. The idea to be conveyed is that similar data points will occupy positions which are closer to each other in the high dimensional space. The random forest classifier is an ensemble learning technique which uses multiple weak learners, which in this case are decision trees to come to a consensus about the output, which in this case is the output label fitted by the clustering algorithm.

In a way, we can view it as a way to understand the similarities between the features, which the k-means clustering exploits to decide the closest centroid. For a description about the k-means clustering algorithm and the random forest classifier, please refer to the corresponding subsections in the annexure.

After determining the weight vectors, the most energetic features are chosen, top 20 according to the decreasing order of their weights. The linear combination is computed only using these 20 features, as they are relevant to our local problem. Note that for each county, the above computation is conducted, and the top 20 features chosen are not the same every time.

Each of the features thus selected demonstrate characteristics specific to the county. The list of the features and the frequency of their occurrence for the Los Angeles county is shown in the Annexure section. Analysing those features we try to find the most common features among the pool of features. This provides us with an important conclusion about the function that we assumed in the beginning. The factors we considered for the estimation of the function are :

1. Population Density (Rank 1 in the table)
2. Employment Status (Rank 7 in the table)
3. Median Household income (Rank 12 in the table)

We can conclude that our initial function $WE = PD \times MHI \times (1 - \%not\ earning)$ was reasonable as all the factors are positively correlated and are among the top influential factors.

5 RESULTS

From the observation in the visualization section we assign weights to the respective features. The weights are crucial in assigning the score of each zip-code in the data-set. We consider the zip-codes as the building blocks and then sum them up county-wise to form the strategic advantage.

5.1 INFERENCES FROM VISUALIZATION

The computation of weights were done by normalizing the correlation we obtained from the function against the feature under consideration. The weights are discussed in the global model section.

$$y_{global} = w_{norm}^T X \tag{5}$$

where,

$$w = [0.66\ 0.077\ 0.14\ 0.09\ 0.09\ 0.10\ 0.11\ 0.08\ 0.18\ 0.05\ 0.18\ 0.05\ 0.09]^T \tag{6}$$

The X vector is the entire feature vector being taken into consideration,

1. Population density
2. Travel Cab, motocycle, public (%)
3. Median Home Value
4. Employment (full + part time %)
5. Single Bedroom rent (in %)
6. Single (in %)
7. Single with roommate (in %)
8. Children (3-17) in private school (in %)
9. Education (More than associate degree in %)
10. Employment Status

Using the above feature vector, we formulate our updated objective function using the weights. Each zip-code is rated according to this weight vector which is to be used later for ATM network wise labeling.

5.2 SCORING

The task involves adding a name-tag classification for each ATM as discussed in the data pre-processing stages. We calculated the average score for each ATM network in each unique zip-code and the frequency of the ATM network in the zip-code according to the equation.

$$S_{atm} = S_{zip-code} \times (Average\ Score\ of\ ATM\ network) \times Frequency \tag{7}$$

where,
S_{atm} = Score of ATM network in a particular zip-code
$S_{zip-code}$ = score of the zip-code
The score of an ATM network in a given county,

$$S_{atm\ (global)} = \sum_{zip \in county} S_{zip} \times (Average\ Score\ of\ ATM\ network|_{zip}) \times Frequency|_{zip} \tag{8}$$

The score we obtain from the above equation is the result from the global model. In the local model, the data is divided into 55 counties. In each county 7 clusters are initially set in order to visualize the important patterns in the data. From the clustered data we reverse engineer the features using the Random Forest classifier. The most important 20 features are chosen per count and weighted sum is computed.

$$S_{atm\ (local)} = W_{rf}^{T} \times x \times (Average\ Score\ of\ ATM\ network|_{county}) \tag{9}$$

where,
W_{rf} = Weight vector obtained from random forest classifier
x_i = Feature data
In this method we directly compute the score of the county. The name-tag weightage of each bank network is computed per county and multiplied with the county score in order to get the ATM network score in the given county. The score of each bank gives us an overview of the relative revenue advantage of an ATM network.

Finally, to compute the overall score, we compute a linear fusion of both global and local features S_{total}:

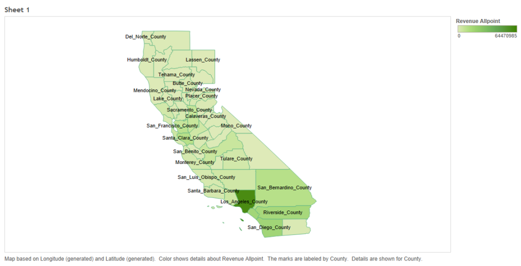
$$S_{atm} = (1-\alpha) S_{county\ (local)} + \alpha S_{county\ (global)} \tag{10}$$

We assumed a value of $\alpha = 0.35$, as we intuitively expect the local features to play a larger role in determining the overall score.

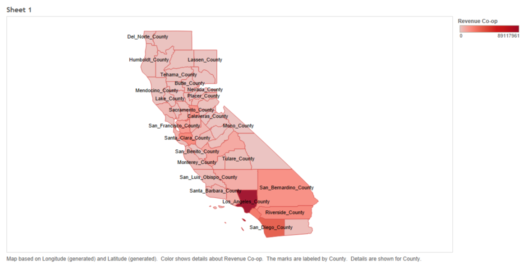
5.3 REVENUE ESTIMATE

We have assumed that an average adult American visits the ATM 60 times per annum (aroung 5 times per month). Also, it is assumed that the transaction cost is \$ 0.5. We approximate the revenue by the following equation (in \$):

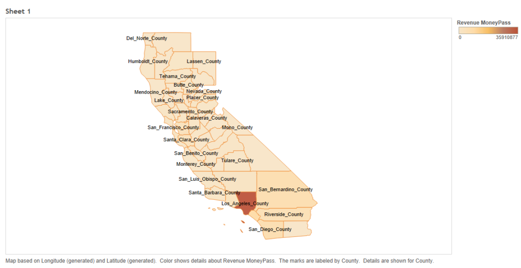
$$Revenue = S_{atm} \times 0.6 \times Population \times 60 \times 0.5 \tag{11}$$



(a) Revenue of Moneypass ATMs by County



(b) Revenue of Allpoint ATMs by County



(c) Revenue of Co-op ATMs by County

Figure 3: Maps denoting revenue of ATM network of the respective bank in each county

6 CONCLUSION

Thus, we would like to propose the above system, and conclude that the hybrid approach is ideologically optimal. The important aspect that the model takes into consideration is that the feature weights varies with county location. We observe that the factors with higher weights vary greatly with the counties. So it would not be very prudent to generalize over all the counties with a specific set of factors. That is why taking the weighted score summation of two models — one which considers local factors and the other which considers the global trends would be the most generalized way to go about solving this problem.

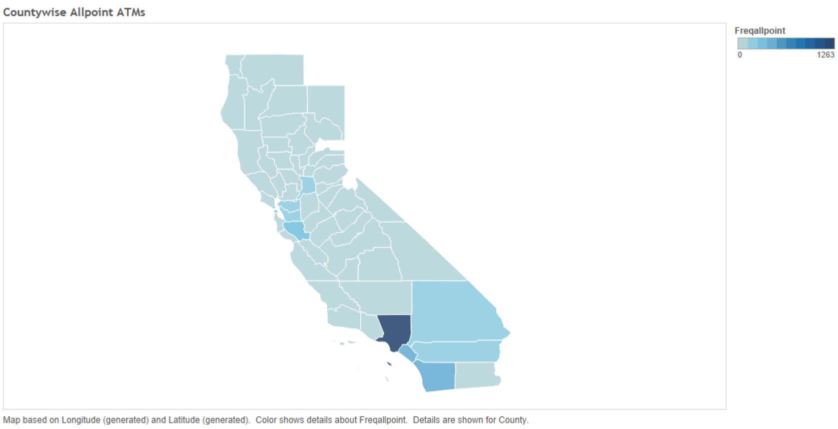
The reason why we took into account both the global and local models instead of considering only the local model is that it would otherwise lead to overfitting, and on the other hand, without local data the model would prove to be very generic. In order to generalize over the entire California region both the models should be accounted for. The weight of the local region in a county is kept higher than the global weight as the significant local factor has to be given more importance. The weight of local model is to be kept as 0.65 and 0.35 assigned to the global model.

From the score estimate of the county regions, we can maximize our revenue if there is a venture for opening a new ATM network. As we already have the score of individual revenues county wise. We can keep in the cost of setting up a new ATM in a particular region as a limiting factor. Then we can optimize the function of the reward (taken as county score) and the penalty (cost of setting up an ATM) over a few counties. Then we can decide which public spots we can set up ATMs in order to reach maximum revenue.

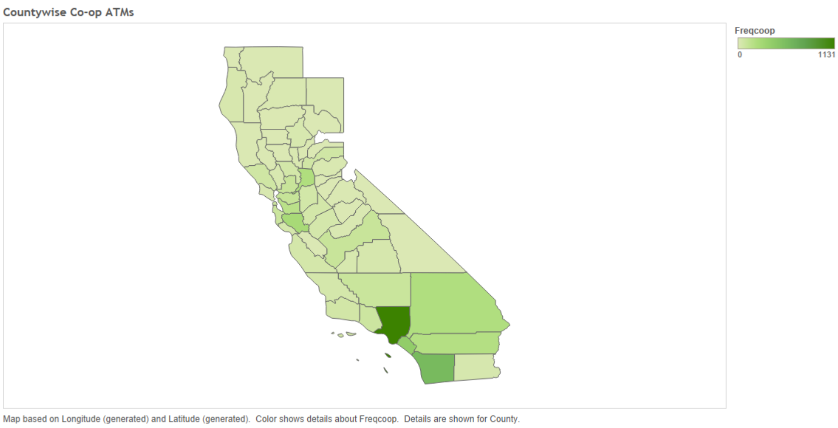
Annexure

7 ANNEXURE

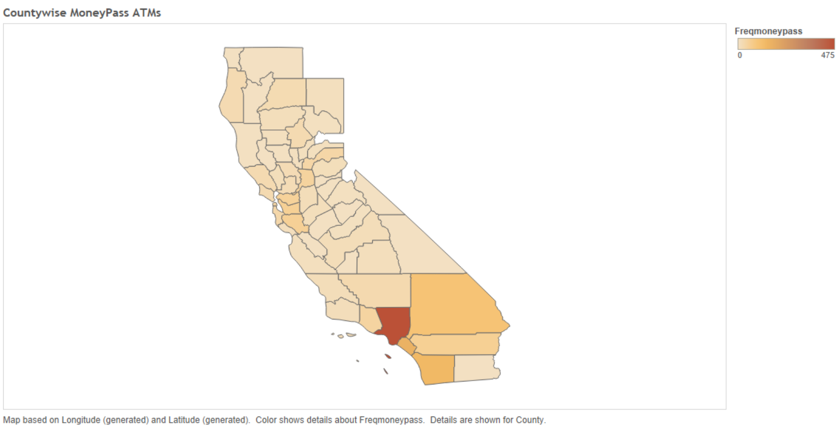
7.1 HEAT MAPS



(a) Allpoint Atms Heat Map



(b) Coop Atms Heat Map



(c) MoneyPass Atms Heat Map

Figure 4: Heat maps of the three atm networks across different counties in CA

7.2 COUNTY-WISE FEATURE WEIGHTS

By using K-means clustering we found out the weights of the feature vector for each individual county. The feature map for Los-Angeles county is show below.

Raw Weights	Features
0.0648538389	Occupied housing units
0.0441433249	Housing units
0.0341489975	Primary school public percent
0.0334325807	Single guardian percent
0.0334284851	Education less than high school percent
0.0315479351	Household occasional use percent
0.0310515523	Household for sale percent
0.028339937	Household vacant other reason percent
0.0282697121	Household rented not occupied percent
0.0281088139	Rental prop 1bedroom percent
0.0277297608	Rental prop 2bedroom percent
0.0267079347	Travel bicycle percent
0.0253925433	Employment parttime percent
0.0235675525	School not enrolled percent
0.0233409342	Husband wife percent
0.0228082624	Population density
0.0220870524	Travel car truck percent
0.021951734	Work at home percent
0.0217348392	Education bachelor percent
0.0217158641	Housing type unoccupied percent
0.0216110645	Singles percent
0.0211646769	Median household income
0.0211460847	Housing occupancy with mortgage percent
0.0207418462	Median home value
0.0192350022	Education high school percent
0.0184424619	Owned household with free clear percent
0.0181347979	Other non institutional percent
0.0176273786	Education master percent
0.0170261506	Nursing facility percent
0.0170029882	Rented occupied households percent
0.0158260468	Employment fulltime percent

7.3 FEATURE LIST

The feature list with frequency is mentioned below.

Feature	Frequency
Population density	27
Occupied housing units	24
Rental prop 1bedroom percent	24
Housing units	23
Rental prop 2bedroom percent	20
Rental prop studio percent	20
Household occasional use percent	20
Household vacant other reason percent	19
Employment part time percent	17
Rental prop 3 bedroom percent	16
Employment fulltime percent	15
Employment unemployed percent	15
School not enrolled percent	13
Household rented not occupied percent	12
Travel bicycle percent	12
Work at home percent	11
Household with kids percent	11
Education professional percent	10
Education high school percent	10
Primary school public percent	10
Husband wife percent	10
Primary school private percent	10
Household without kids percent	10
Singles percent	10
Housing occupancy with mortgage percent	10
Median home value	10
Education associate degree percent	9
Rented occupied households percent	9
Median household income	9
Housing type unoccupied percent	8
Nursing facility percent	8
Travel motorcycle percent	8

The above table denotes how many times a feature occurs in the top 10 important features derived from the local model. The local models maps the most important features for a particular county.

7.4 K-MEANS CLUSTERING

k-means clustering is a method of vector quantization which is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means because of the k in the name.

Given a set of observations $(x_1, x_2, ..., x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets S = $\{S_1, S_2, ..., S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (sum of distance functions of each point in the cluster to the K center). In other words, its objective is to find:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \tag{12}$$

where $\boldsymbol{\mu}_i$ is the mean of points in S_i.

We have used$k$-means clustering in the local model of our analysis for county wise atm clustering to know the features playing important role in the clustering locally(i.e. countywise) and we have used the value of k=7 in the clustering process.

7.5 RANDOM FORESTS

Random forests or random decision forest is an ensemble learning method for classification, regression, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees habit of over-fitting to their training set.

Random forests can be used to rank the importance of variables in a regression or classification problem in a natural way. The following technique was described in Breiman's original paper and is implemented in the R package randomForest.

The first step in measuring the variable importance in a data set $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^{n}$ is to fit a random forest to the data. During the fitting process the out-of-bag error for each data point is recorded and averaged over the forest (errors on an independent test set can be substituted if bagging is not used during training).

To measure the importance of the j-th feature after training, the values of the j-th feature are permuted among the training data and the out-of-bag error is again computed on this perturbed data set. The importance score for the j-th feature is computed by averaging the difference in out-of-bag error before and after the permutation over all trees. The score is normalized by the standard deviation of these differences.Features which produce large values for this score are ranked as more important than features which produce small values.

7.6 REVENUE GENERATED BY ATM NETWORKS

County	All Point	Money Pass	Co Op
Alameda	819398.944	4387322.23	12826650.83
Amador	199941.0298	58428.88155	159176.0886
Butte	1587575.821	1507162.992	1229689.187
Calaveras	232710.5363	189623.8482	100043.615
Colusa	110983.7576	135796.2424	0
Contra Costa	6720804.364	3908118.481	7871585.155
Del Norte	95979.13028	172762.4349	172762.4344
El Dorado	1313771.598	805297.133	1083221.269
Fresno	6006535.749	1068274.627	8395937.624
Glenn	93838.17555	293485.2908	50976.53366
Humboldt	562650.5446	737446.8871	772314.5683
Imperial	1306172.457	0	2215689.543
Inyo	223037.3022	58122.69781	0
Kern	4032793.398	2028534.694	7399413.907
Kings	817470.8594	99736.24571	1710288.898
Lake	545133.1897	84019.30097	276859.5094
Lassen	227571.8774	152554.5611	193983.5621
Los Angeles	64470985.25	35910877.43	89117961.31
Madera	38433.6534	341137.7226	1281830.624
Marin	1838959.139	1265424.686	1101478.175
Mariposa	0	0	199116
Mendocino	466011.0281	169167.5669	618611.4038
Merced	1425326.782	148224.5828	2452832.639
Mono	172206	0	0
Monterey	2751421.573	335696.0205	3043736.406
Napa	1056665.836	438848.8489	958479.3152
Nevada	629045.6	447001.6389	230176.7611
Orange	19159495.2	9550374.886	24241323.96
Placer	1750306.271 1505071.629	2575182.106	
Plumas	71136.04853	123299.9515	0
Riverside	17512948.57	5782681.626	18669839.8
Sacramento	7251358.125	3166276.336	14720895.54
San$_{Benito}$	605308.131	0	273829.869
San$_{Bernardino}$	11476180.63	7553921.501	16184719.83