

# Story of A Slum: Pamplona Alta



Every year, in partnership with the Solidaridad en Marcha, Catholic churches throughout the city of Lima, Peru deliver thousands of Christmas gift boxes to the poorest residents in the city. The campaign, called La Caja Del Amor, has been in operation for years and has subsequently built long-term and well-sustained relationships with many community leaders in these areas.

The list of gift recipients is built in coordination with the local community leaders, most of whom oversee around 150 families. These leaders choose the 5 to 10 families in most need of assistance in their respective community to receive the gift boxes.



**La caja del amor**  
Un regalo de familia a familia

These networks served as a point of strength when we began to explore the creation of an in-depth survey to better understand the poorest urban populations in Peru. However, our focus needed to also be on action-ability of insight and the concrete opportunity of enacting positive change in the lives those we surveyed.

Therefore, we decided to focus on a region of high need, and an area where Solidaridad en Marcha had a significant footprint. This process of elimination led us to the region of Pamplona Alta near the San Juan de Miraflores municipality of Lima.

# Table of Contents

<b>Pamplona Alta, A Brief History</b>	<b>3</b>
<b>Information Description</b>	<b>5</b>
○ <u>Survey</u>	
○ <u>Dataset</u>	
<b>Understand Population</b>	<b>7</b>
○ <u>Averages</u>	
○ <u>Distribution</u>	
○ <u>Correlation</u>	
<b>Understand Population Within</b>	<b>13</b>
○ <u>Clusters</u>	
○ <u>Communities</u>	
<b>Drill Down Into Specific Issues</b>	<b>16</b>
○ <u>Household Size and Income</u>	
○ <u>Role of Men</u>	
○ <u>Grandparents, Children and Commuters</u>	
○ <u>Females and Finance</u>	
○ <u>Newly Established Households</u>	
<b>Conclusion</b>	<b>23</b>
○ <u>Summarized Solutions and Potential Impact</u>	



## Pamplona Alta: A Brief History

Pamplona Alta is a shanty town riddled with extreme poverty, it has an absence of infrastructure, and a large portion of its community is without access to many basic human needs. The region has no public works, no paved roads, no public electricity, nor public access to water and sewage. If these services exist at all, they are provided by private companies at a premium price.

Water trucks provide the only (non-drinkable) water and they often cannot access many of the homes, especially in the upper portions of the region. Shallow latrines serve as a poor substitute for sewage and along with the local pig farms, contribute to a high rate of parasitic infections, particularly among children.

The area was first populated in the 1990's as a result of a decade of terrorism that affected the entire country of Peru. During that period, populations from the surrounding regions began a mass migration to the outskirts of Lima, leaving the lives they knew to find safety in the proximity of the city. Many only spoke Quechua, few had employment for years after arrival, and none of these new residents owned the land on the edge of Lima where they would now call home.



It has predominantly been this issue of land ownership that has ensured the continued impoverishment of the people of Pamplona Alta. Why land ownership? Well, owning the land is necessary, according to the government of Peru, for those in Pamplona Alta to receive basic municipal services. In order to attain roads, plumbing, water, schools and many other basic human needs, they must first own the land on which their homes sit.

The land is rocky, mountainous and un-arable, yet to the people of Pamplona Alta, it is home. Inside of this difficult landscape, there are two prevailing realities: those who squatted on public land and those who squatted on private land.



The valleys of this mountainous region were owned predominantly by pig farmers upon arrival, and even today many families live side by side to pigs in the lower part of Pamplona Alta. The owners of the land have lacked the resources to forcibly evict their unwanted tenants since their arrival, yet these tenants have now called Pamplona Alta home for over 20 years and despite their best hopes, still, have a nearly impossible chance of owning the land on which their homes sit. As a result, many of these structures lack stewardship, the inability to own does not reasonably warrant investment and in turn, the people there exist in squalor.

The story of those in the upper mountain portions of Pamplona Alta is a different, yet only a slightly less dire narrative. For people who have built their homes on government-owned land, ownership is possible, however, only after a laundry list of nearly impossible requirements, given by the government of Peru, is achieved. Land ownership is possible to those who squatted on government land if they have:

- been there for longer than 5 years
- access to water
- access to electricity
- safe access to home
- community centers/parks within close proximity of their house

In practice, the government of Peru is asking the poorest and least advantaged people in their country to not only sustain life on a few dollars a day, but also build a road to create access for a privately priced water truck to get to their home, to pay to bring privately priced electricity to their home, to build community centers and to construct safe access to their homes from the bottom of the mountain.



This task list is absolutely impossible on their own, and organizations like Solidaridad en Marcha have helped make home ownership a reality to some of the people of Pamplona Alta, however, this battle is still uphill, and full of rocks and mud.

It is important to keep in mind that, yes, the government has been far from perfect, however, they are also faced with an incredible dilemma. The government entities feel obligated to protect land rights, discourage further squatting on un-owned land and legally address how to approach a population of 350,000 people already living on land that is not their own. There are massive policy improvements that can be made, but there is no denying the incredible legal, economic and political complications involved.



All things considered, the subsequent reality for the people of Pamplona Alta is that their children are frequently sick, their jobs are too far away, under-education is inevitable, many families are broken and the lack of government support ensures the existence of a dark economy, thus extending their lack of access to credit and upward mobility.

Deeper than these economic indicators of poverty, the people of Pamplona Alta suffer from the poverty of dignity. Many members refer to themselves as “the forgotten ones,” election promises come and go without much change and through the confusing red tape and legalities, they still find themselves without access to basic human needs. In spite of that, the people of Pamplona Alta are resilient, hard-working and deeply committed to building the best life they can for their families.

## Information Description: Survey

Understanding the many facets of those in Pamplona Alta was integral in the way we built our survey, the questions we asked, and the way we asked them. We raised some questions that we heuristically had an intuitive idea of the answer, but needed to understand the severity. Yet, others we asked in order to gain insight into the tools we may have available to us within our solution set.

Lastly, we inquired of economic indicators, religious factors and family structure. All intended to paint a picture of the lives of those in Pamplona Alta and to possibly determine causality between the various characteristics.

Using this form, we were able to fully survey over 500 families and after extracting all personally identifiable data, we built a dataset that held great potential for a better understanding of the lives of those in Pamplona Alta and the possible routes available to help them.

Número de la familia _____			 <b>La caja del amor</b> Un regalo de familia a familia			 <b>Solidaridad en Marcha</b>			AAHH _____ Encargado _____ Celular _____		
	NOMBRE	APELLIDOS	RELACIÓN	NOMBRE	SEXO (F/M)	EDAD					
MADRE	_____	_____	_____	_____	_____	_____					
PADRE	_____	_____	_____	_____	_____	_____					
DIRECCIÓN Mz. _____ Lt. _____			_____	_____	_____	_____					
			_____	_____	_____	_____					

**Marca una 'x' según sea "Sí" o "No"**

¿Tu teléfono tiene internet?	Sí	No
¿Puede la cisterna de agua llegar a tu casa?	Sí	No
¿Tienes una cuenta en el banco?	Sí	No
¿Vas al menos una vez al mes a la iglesia?	Sí	No
¿Sueles dejar a tus hijos solos en casa?	Sí	No

**Responde solo con números**

¿Cuántas personas viven en tu casa?	Número _____
¿Hace cuánto tiempo vives en esta casa?	_____
¿A qué edad tuviste tu primer hijo?	_____
¿Cuántas personas trabajan en tu casa?	_____
¿Cuánto tiempo te toma llegar a tu trabajo?	_____
¿Cuántos días de colegio pierden tus hijos al mes?	_____
¿Cuál es el ingreso mensual de tu familia?	_____
¿Cuántas personas en tu casa son bautizadas?	_____



## Information Description: Dataset

Months of conversations, meetings, reading and collaboration with community members came into the building of this survey. Qualitative analysis helped us produce a dataset with the potential of performing multiple quantitative analyses.

The completed dataset built from the original survey contains 21 household characteristics (variables) of 507 families across 81 communities, providing information on a total of 2,231 people for us to explore, visualize and perform analysis on. The complete dataset can be found on my GitHub account, [here](#).

1. **fam\_n** – *numeric factor* – Unique identifier for each family.
2. **internet** – *binary* – Does your phone have internet?
3. **agua** – *binary* – Can the water truck get to your house?
4. **banco** – *binary* – Do you have a bank account?
5. **iglesia** – *binary* – Do you go to church at least once a month?
6. **dejar\_hijos** – *binary* – Do you leave your children home alone (when you go to work)?
7. **cuantas\_personas** – *numeric* – How many people live in your house?
8. **tiempo\_casa** – *numeric* – How long have you lived in your house?
9. **primer\_hijo** – *numeric* – At what age did you have your first child?
10. **cuantas\_trabajan** – *numeric* – How many people in your house work?
11. **tiempo\_trabajan** – *numeric* – How long does it take to get to your job?
12. **pierden\_colegio** – *numeric* – How many days a month do your children miss school?
13. **ingreso** – *numeric* – What is your monthly household income?
14. **bautizadas** – *numeric* – How many people in your family are baptized?
15. **direccion** – *character factor* – Name of neighborhood.
16. **padre** – *binary* – Does the father of the children live in the home?
17. **madre** – *binary* – Does the mother of the children live in the home?
18. **F** – *numeric* – Count of females in the home.
19. **M** – *numeric* – Count of males in the home.
20. **niños** – *numeric* – Count of children 18 and younger in the home.
21. **mayores** – *numeric* – Count of adults 65 and older in the home.

**Reminder:** I will at times use technical language, but I encourage you to keep reading through, because I will also accompany every technical insight with an explanation in simple language, relevant to the question at hand. Also note, for binary variables, 1 is yes and 0 is no.

## Understand Population: Averages

We can see that a small proportion of those surveyed had bank accounts and internet access, while a large proportion went to church regularly and a majority had access to drinkable water though not overwhelmingly so. In terms of family dynamics, we see that only half of fathers are present in the home and most homes have a mother present. The initial pulse we get from the binary proportional averages is that we can affirm some of our pre-conceived ideas surrounding broken families and presence of a dark economy, implying sparse routes to traditional credit sources.

Have Internet Access 11.06%	Have Access to Water 70.00%	Have Bank Accounts 3.89%
Attend Church Services 88.71%	Have Child Care 57.45%	Father Present in Home 54.82 %

Inspecting the numeric means of our survey results, we can begin to get a sense of the lives lead by the poorest of the poor in Pamplona Alta. We see that the average time in the home is nearly triple the required 5-year requirement of the government for ownership. However, we also see that the average household of 5 is usually living off of the salary of a single person, and that average value is only 649 soles a month (\$6.30 USD a day per household or \$1.26 per person). Looking to the religious aspects of the survey, it is also interesting to note that on average only 2/5 members of households are baptized though a majority attend church services regularly.

Household Size 5 people	Time in Home 14 years	Mother's Age at First Child 21 years old	Work Commute 1 hour 10 minutes
Working in House 1 to 2 people	School Missed per Month 2 days	Household Income per Month 649 Soles	Baptized in Household 3 people
Women in House 3	Men in House 2	Children in House 3	Elderly in House 0 to 1

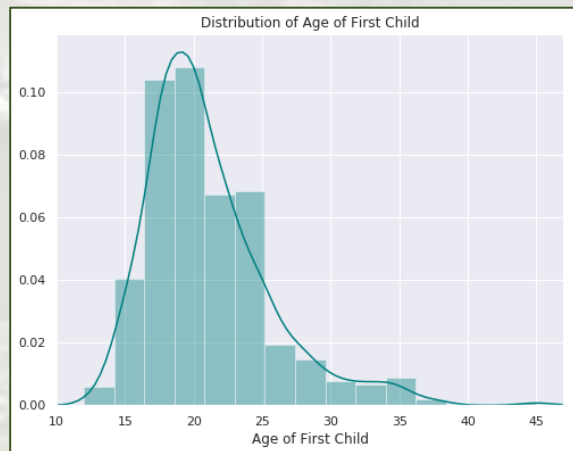
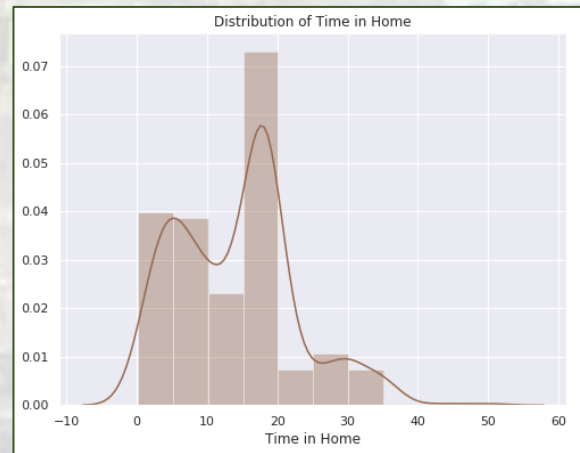
We do not want to make any extraneous assumptions from these average values, nevertheless, these initial figures have allowed us to paint a faint picture of the lives led by those surveyed and provided a question set for us to move forward in our analysis.



## Understand Population: Distribution

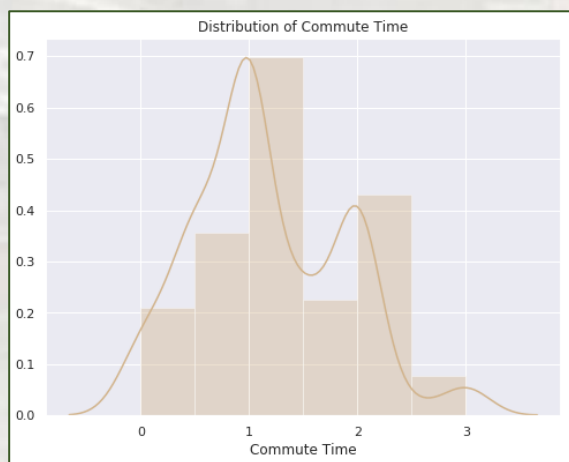
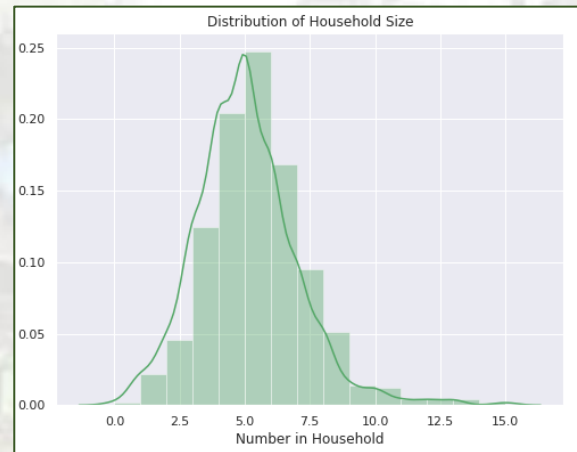
Average values can give us a glimpse into the character of a variable, however, the distribution can tell us even more. Given the above results I wanted to take the time to look into certain variables of intrigue. This process gives shape to the values within, as well as helps us spot outliers that may significantly affect our future modeling.

We can see that distribution of time-in-home is tri-modal, meaning that there are three time groupings in which people have been in their homes: many have been there less than 10 years, some more than 25, but most have been in their homes between 15-20 years. This is an encouraging data point when considering the government requirements for home-ownership.



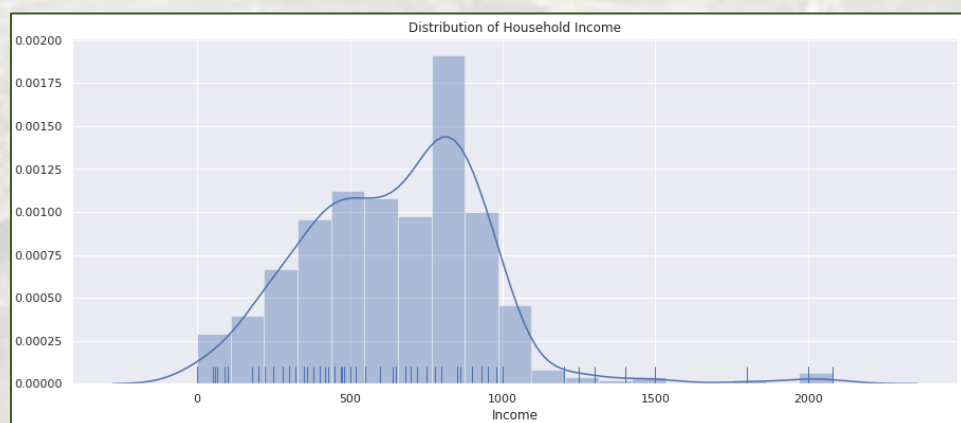
The second distribution, is that of the first child. The average value for this number was 21, which given our knowledge of the region, felt higher than we had experienced. The distribution confirms the fact that a majority of the women are having their first child at or before the age of 20. The average value was being pulled upwards due to the positive skew of the distribution. In Pamplona Alta, the reality of having children before the age of 17 is often dropping out of primary school, therefore, seeing fewer of these than we anticipated is mildly encouraging.

The third distribution of note is household size, the values are, more or less, normally distributed, with a majority of the households being between 4 and 6 people per household. We can see that there are some houses with as many as 15, but that number is small enough that those outliers should hold little weight to our predictive modeling further on.



Next, looking to commute time, we see a mostly bi-modal distribution, with a majority of people traveling just over an hour to work, and another traveling over two hours to work. This confirms the difficulty we've seen for people in Pamplona Alta to find employment near to home.

When attending to the distribution of household income below, we decided to remove two outliers early in the exploratory process. Nevertheless, we see that a majority of the families surveyed had a household income between 500/S and 1000/S Soles per month (\$147.84 to \$295.68 USD). Most estimates for income per capita in Peru are mostly north of 1500/S per month, therefore we can affirm that our survey truly touched the poorest of the poor.





## Understand Population: Correlation

Beyond univariate analysis, the study of only one variable, we want to start looking into how different factors behave in comparison to one another and to do so, we will initially use correlation. Correlation is a mutual relationship, or connection between two or more characteristics/variables, therefore when comparing numeric relationships with correlation, the range is from -1 to +1.

If two variables are perfectly correlated, they will have a correlation of 1, while if they are perfectly opposite, they will have a correlation of -1. If there is no correlation, they will have a 0 value for correlation. In the chart below contains every variable we have available for comparison, and matching the values across the X axis with those on the Y, we can see the correlation value between those two values.

internet	1.00	0.04	0.04	-0.14	0.05	0.03	-0.06	-0.02	0.09	0.01	0.01	0.24	-0.01	0.11	-0.01	-0.03	0.05	0.00	-0.05
agua	0.04	1.00	0.05	0.00	-0.03	0.09	0.08	0.02	-0.03	-0.04	0.12	0.00	0.06	0.03	0.07	0.06	0.09	0.11	0.08
banco	0.04	0.05	1.00	-0.12	0.11	0.07	0.00	-0.08	0.09	0.04	0.04	0.20	0.05	-0.06	-0.09	0.09	0.02	0.03	-0.05
iglesia	-0.14	0.00	-0.12	1.00	-0.05	0.04	0.07	0.13	-0.08	-0.01	-0.03	-0.06	0.07	0.04	-0.03	0.01	0.04	0.05	0.05
dejar_hijos	0.05	-0.03	0.11	-0.05	1.00	0.06	-0.06	-0.02	0.13	-0.02	0.13	-0.03	-0.05	-0.13	0.03	0.12	0.02	0.19	0.01
cuantas_personas	0.03	0.09	0.07	0.04	0.06	1.00	0.18	-0.08	0.22	-0.00	0.02	0.28	0.29	0.22	0.10	0.53	0.47	0.68	-0.04
tiempo_casa	-0.06	0.08	0.00	0.07	-0.06	0.18	1.00	0.06	0.04	-0.07	-0.07	0.03	0.28	-0.07	-0.01	-0.01	0.03	-0.02	-0.02
primer_hijo	-0.02	0.02	-0.08	0.13	-0.02	-0.08	0.06	1.00	-0.06	0.04	0.06	0.06	-0.01	0.09	-0.08	-0.10	-0.04	-0.15	0.06
cuantas_trabajan	0.09	-0.03	0.09	-0.08	0.13	0.22	0.04	-0.06	1.00	0.11	0.19	0.19	0.12	0.09	0.00	0.13	0.08	0.08	0.01
tiempo_trabajan	0.01	-0.04	0.04	-0.01	-0.02	-0.00	-0.07	0.04	0.11	1.00	0.01	0.12	0.05	0.02	-0.06	-0.00	0.02	-0.02	0.00
pierden_colegio	0.01	0.12	0.04	-0.03	0.13	0.02	-0.07	0.06	0.19	0.01	1.00	-0.10	-0.04	0.06	0.00	0.01	0.09	0.09	-0.02
ingreso	0.24	0.00	0.20	-0.06	-0.03	0.28	0.03	0.06	0.19	0.12	-0.10	1.00	0.13	0.25	-0.08	0.10	0.19	0.11	-0.02
bautizadas	-0.01	0.06	0.05	0.07	-0.05	0.29	0.28	-0.01	0.12	0.05	-0.04	0.13	1.00	0.01	-0.04	0.04	0.15	0.12	0.04
PADRE	0.11	0.03	-0.06	0.04	-0.13	0.22	-0.07	0.09	0.09	0.02	0.06	0.25	0.01	1.00	-0.17	0.03	0.39	0.07	-0.03
MADRE	-0.01	0.07	-0.09	-0.03	0.03	0.10	-0.01	-0.08	0.00	-0.06	0.00	-0.08	-0.04	-0.17	1.00	0.16	-0.03	0.09	0.04
F	-0.03	0.06	0.09	0.01	0.12	0.53	-0.01	-0.10	0.13	-0.00	0.01	0.10	0.04	0.03	0.16	1.00	-0.11	0.59	0.04
M	0.05	0.09	0.02	0.04	0.02	0.47	0.03	-0.04	0.08	0.02	0.09	0.19	0.15	0.39	-0.03	-0.11	1.00	0.54	0.02
NINOS	0.00	0.11	0.03	0.05	0.19	0.68	-0.02	-0.15	0.08	-0.02	0.09	0.11	0.12	0.07	0.09	0.59	0.54	1.00	-0.10
MAYORES	-0.05	0.08	-0.05	0.05	0.01	-0.04	-0.02	0.06	0.01	0.00	-0.02	-0.02	0.04	-0.03	0.04	0.04	0.02	-0.10	1.00
	internet	agua	banco	iglesia	dejar_hijos	cuantas_personas	tiempo_casa	primer_hijo	cuantas_trabajan	tiempo_trabajan	pierden_colegio	ingreso	bautizadas	PADRE	MADRE	F	M	NINOS	MAYORES

Initially, the correlation strengths are discouraging, we are seeing largely weak correlations between most of the variables, meaning that the predictive power of models we will build could be difficult to come by. However, we see some relationships with more than weak positive correlations, so let's dig into a few of those:

- The more time someone has been in their house, the higher the likelihood that they will be baptized. Thus, implying those who are established in a community are also established more in the local church, while newly established / less stable families have a more difficult time attaining access to the sacraments.
- If the father of the household is present, there is not necessarily more workers in the house, but the household size and income are greater when he is present. What this could mean is that having a father present allows the family to make more money, thus the ability to support a larger household size.
- The relationship between internet and income is one of the stronger relationships, however the correlation does not mean causality. It is assumed that one needs money to pay for internet, however, it could be interesting to measure the causality of this relationship in order to see if access to internet also has the ability to propel the household income higher after attaining it. This is the same case for bank accounts, insofar as people with the most income have bank accounts, but with correlation we do not know if bank accounts are causing higher income; for this, other analysis will be needed.
- We see that the younger the mother was when she had her first child, the more children she ends up having in the household, also those who became mothers earlier are less likely to be involved in the local church.
- The larger the household size, the longer time in the house. This could just be a proxy for time, however, it could also suggest that the larger the family the more established they are in the community and all the subsequent externalities associated with that.

It is important to note the strength of these hypotheses are only as strong as the underlying correlations behind them, however the exercise above gives us an opportunity to move forward with some understanding and ask better questions as we go on.



## Understand Population Within: Clusters

The survey delivered 21 characteristics about every household and using these characteristics, we have the potential to discover natural groups that exist within the total population. In order to bring these groups to light, we use what is called k-means cluster analysis. This technique places our 21 personal characteristics into a 21-dimensional vector space, and then determines centroids based on a given number of groups. In laymen's terms, this means that k-means finds the people that are similar to one another.

Understanding what cluster (group) a household may fall into could help identify the optimal aid needs of that household, as well as give shape to the community itself. We discovered 4 naturally occurring groups within the population and through the average values of the characteristics, we were able to discover a small story of the lives they lead.

group_name	internet	agua	banco	iglesia	dejar_hijos	cuantas_personas	tiempo_casa	primer_hijo	cuantas_trabajan	tiempo_trabaja	pierden_colegio	ingreso	bautizadas	PADRE	MADRE	F	M	NINOS	MAYORES	cluster_size
Balanced and Stable	15.7%	67.5%	3.1%	90.6%	38.7%	5.628	13.073	21.146	1.340	1.245	1.636	PEN 834.50	51.1%	69.1%	95.8%	3	2	3	0.042	191
Small and Vulnerable	6.6%	73.5%	2.2%	87.5%	49.3%	4.574	14.283	21.478	1.282	1.098	2.268	PEN 508.20	53.5%	46.3%	97.8%	2	2	3	0.074	136
Better off, but not spiritually	25.0%	83.3%	33.3%	66.7%	58.3%	7.500	16.108	20.083	2.167	1.188	2.388	PEN 1,485.83	51.3%	66.7%	100.0%	3	3	4	0.000	12
Poorest, but not in spirit	1.6%	67.2%	1.6%	90.2%	39.3%	5.016	14.180	19.574	1.086	1.084	2.144	PEN 217.30	58.8%	34.4%	98.4%	3	2	3	0.033	61

### Group 1: Poor, Balanced and More Stable

- This group has the most balanced characteristics of those surveyed. 15% have internet, they have high rates of church attendance, but low rates of baptism. They have to travel far to work, but their households are more balanced, with a high rate of fathers in the home. They are the most likely to have grandparents in the household, thus leave their kids at home less, which quassai validates our thoughts on grandparents being great options for babysitters.

### Group 2: Small and Vulnerable

- This group is interesting, in that they earn far less, fathers are still in only 43% of homes, and all other positive statistics are muted by this. Their families are the smallest, but they also have the highest likelihood of having a grandparent in the household.

### Group 3: Better off, but not spiritually

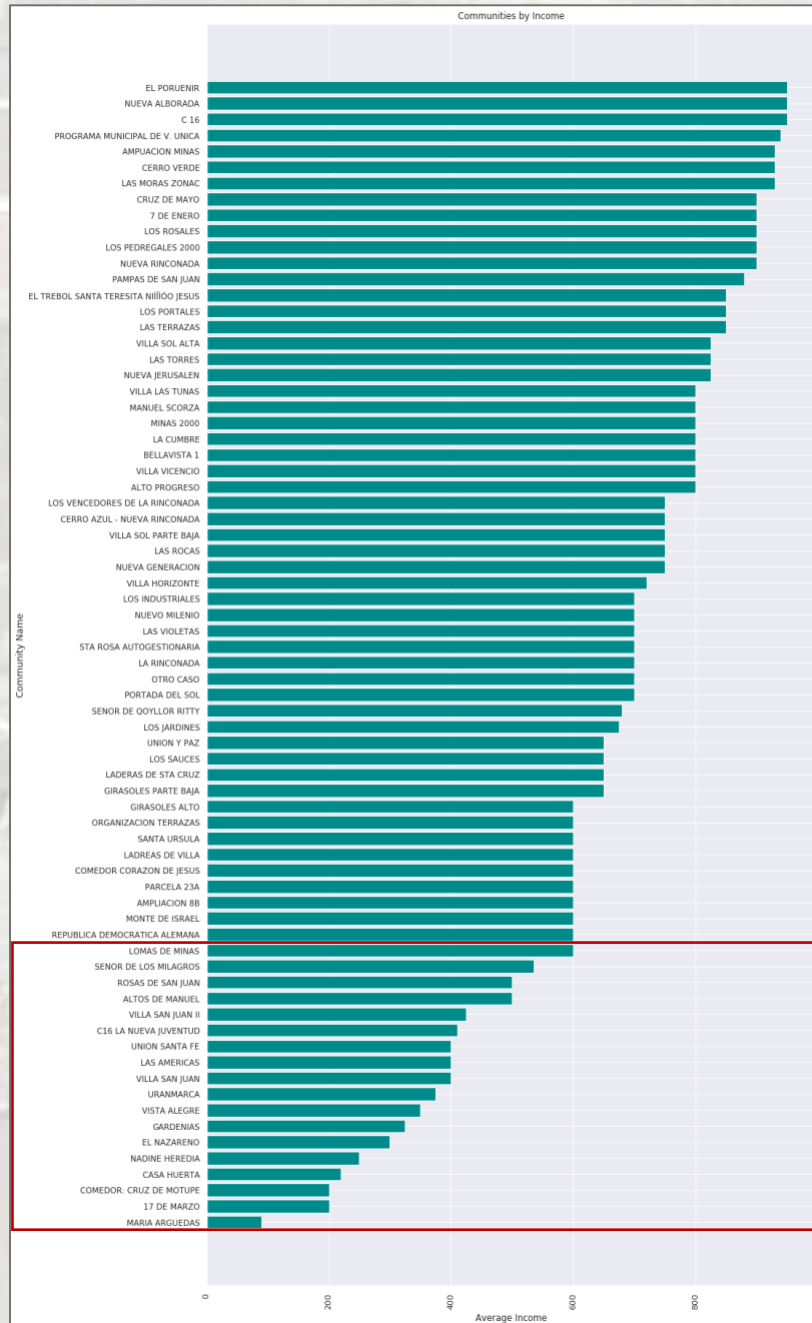
- Wealthiest group by far, with the highest rates of water, bank accounts, time in house. They had their first child later and have to travel farther to work. However, there is a spiritual gap with this group. We see that they have lower rates of baptism and go to church the least of all groups.

### Group 4: Poorest, but not in spirit

- This group is the poorest among the groups, but has much higher baptism and church rates. Only 67% have water access, and 1.6% have bank accounts or internet. They do not leave their kids at home and they have short commutes to work. The only 34% of households have fathers and the families are smaller as a result. They are however established in the community and have been there for a longer period.

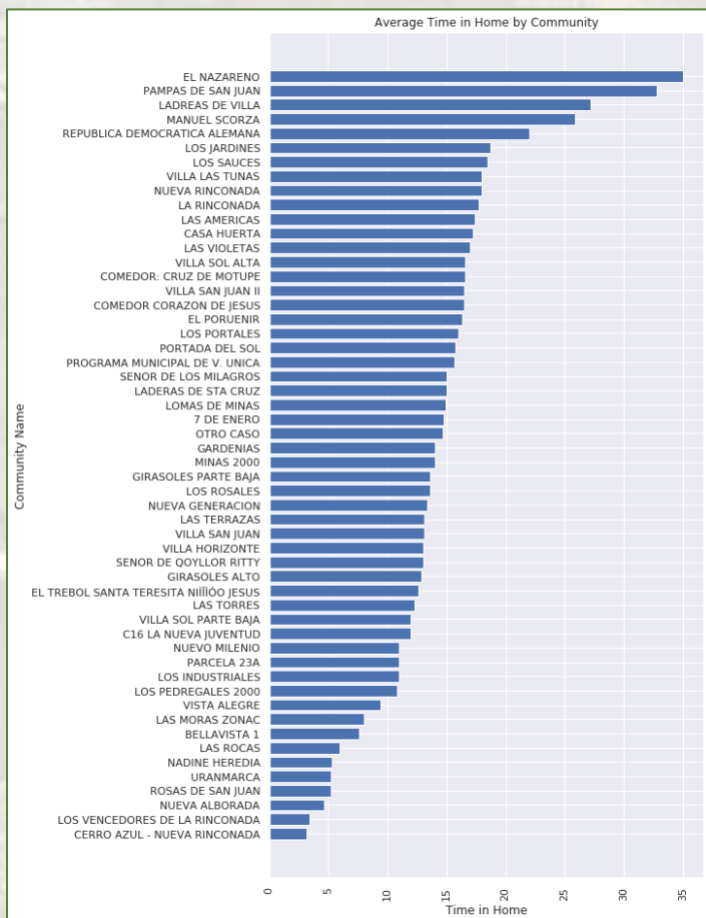
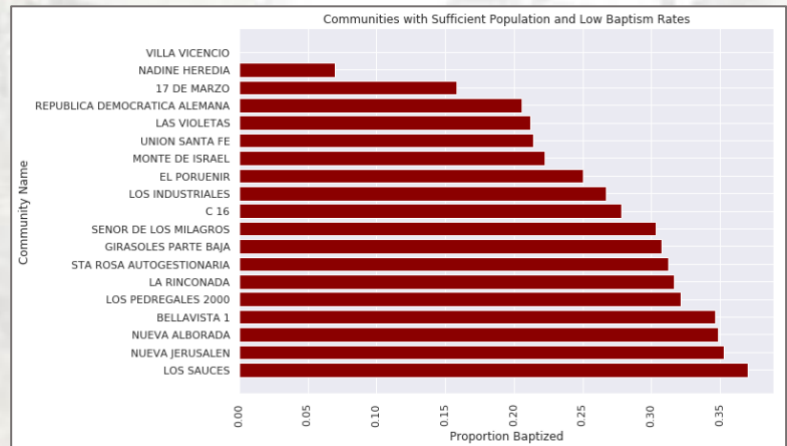
## Understand Population Within: Communities

The survey reached 81 of the smaller communities within Pamplona Alta, and as discussed in the introductory story, we see that reality is largely varied amongst the communities within. In order to get a general idea of the state of communities, we have the average income by communities with more than 10 individuals. Similar to the income distribution charts, we can see that most communities have and an average household income between 600 and 800 soles a month. There is, unfortunately, a very poor set of communities that earn far less than 600 soles a month.





This graph displays the communities with the lowest quartile of baptism rates amongst all communities with more than 10 individuals represented. This list of communities could be given priority for catechesis efforts by various Catholic groups that come through.



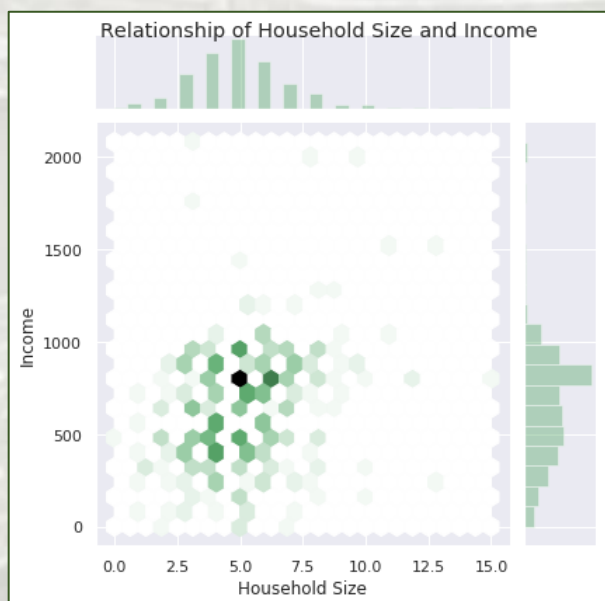
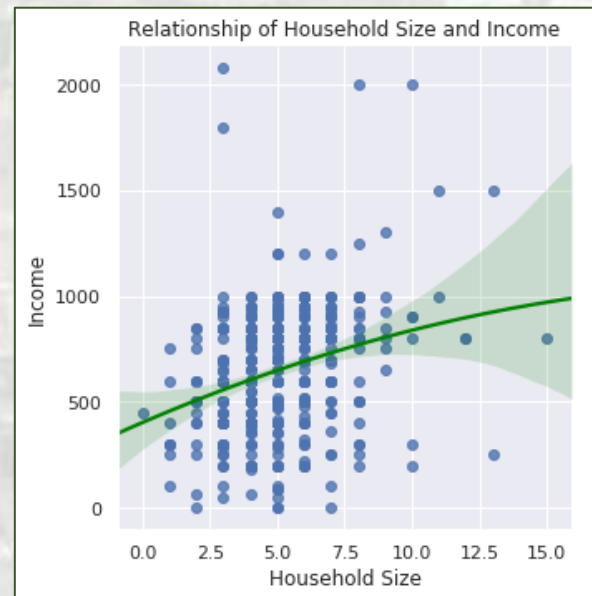
The time someone has been in their home is an incredibly powerful figure that impacts home ownership, community involvement, water access and general stability of a family. Given the nature in which families settle in Pamplona Alta, those communities whom have been in their home less, have access to less services, but are not particularly poorer (lower income) than other communities. This is evident when looking at the El Nazareno community, which has been in their homes the longest, but has one of the lowest average incomes amongst communities represented.

There are many questions that can be asked of metrics partitioned by community. We have given access to a raw csv of these averages and they are located [here](#).

## Drill Down: Household Size and Income

The correlation gives us a high level understanding as to how two variables behave with one another, but the nature of that relationship is much deeper than correlation. Regression attempts to measure the relationship between a dependent variable and a predictor (independent). After, modeling multiple relationships using simple regression techniques, we determined the following points to be of significance to our understanding of the population:

As the household size gets bigger, there is a slight increase in household income, however, this relationship begins to flatten out after a household size of 7. Thus, suggesting that families between 5 and 7 people, on the whole, have more financial strength than those with small or large families. In regards to aid work, this could mean that we need to give special attention to these big and small households.



This visualization measures the same two variables as above, but instead of mapping a line representing the relationship, it visualizes density. We can see the densities along the right and top sides of the graph, and the hex shapes in the middle represent where these densities overlap. Therefore, what this graph is implying is that a majority of those surveyed have a household size of 5 with a household income of 780, offering us much more of a picture the the reality of this relationship.



## Drill Down: Role of Men

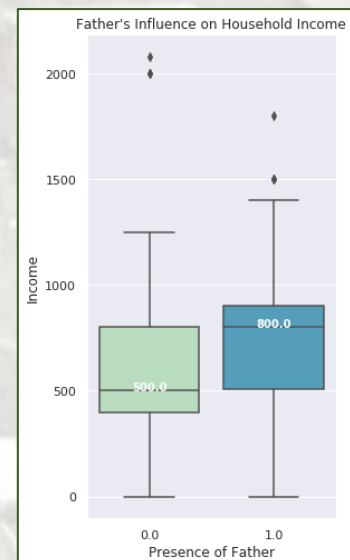
Studying correlations we continued to notice the comparatively large positive relationship between the presence of the father in the household with other variables in our dataset. To understand this further we performed a logistic regression, with a particular focus on the coefficients, which gave us insight into the relationship between the presence of the father and each particular variable.

At 84% accuracy the logistic regression results were fairly accurate, but the key for us is the coefficient values. In the table, all the variables involving household size have positive coefficients with the presence of a father at home. Income also has a positive coefficient with the father's presence, confirming previous insight suggesting that presence of the father provides a significant increase on household income.

M	2.324768
NINOS	-2.254182
F	1.694558
tiempo_casa	-0.399569
MADRE	-0.362897
ingreso	0.331988
MAYORES	-0.266467
cuantas_personas	0.245473

	coef
internet	168.9449
agua	-9.1563
banco	285.9639
iglesia	-25.1677
dejar_hijos	-6.8779
cuantas_personas	34.0702
tiempo_casa	-0.4739
primer_hijo	8.5296
cuantas_trabajan	46.5471
tiempo_trabajan	44.6666
pierden_colegio	-16.1542
bautizadas	39.6487
PADRE	88.1163

We also performed a multiple linear regression in order to predict income of each household solely using the other survey questions. We were able to explain 87% of the variance in that model, however, like the logistic regression, the coefficients provided the most insight. We can see that the presence of the father has a significant positive coefficient in household income.

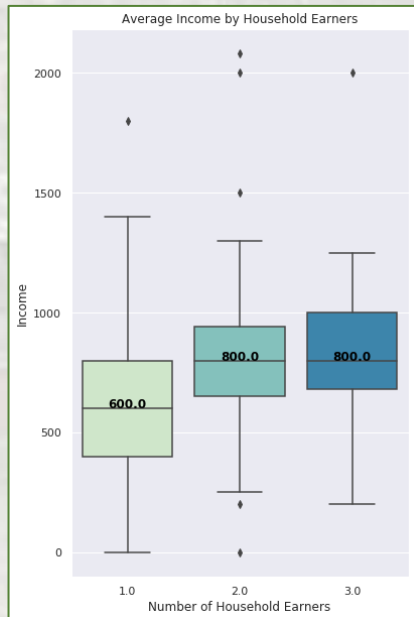


The above coefficients are confirmed further by the box and whisker visualization of income and presence of the father in the household. We can see that having a father in the home results in an average of S/300 more soles a month than households without fathers present.

The value of fathers in the household is validated by multiple analyses, however only 54% of households in our survey have a father present. They are better top-line earners and provide significant strength to the family income in a way that allows them to also support bigger families. The trick will be to see how we can support them to stay home and be present for their families.

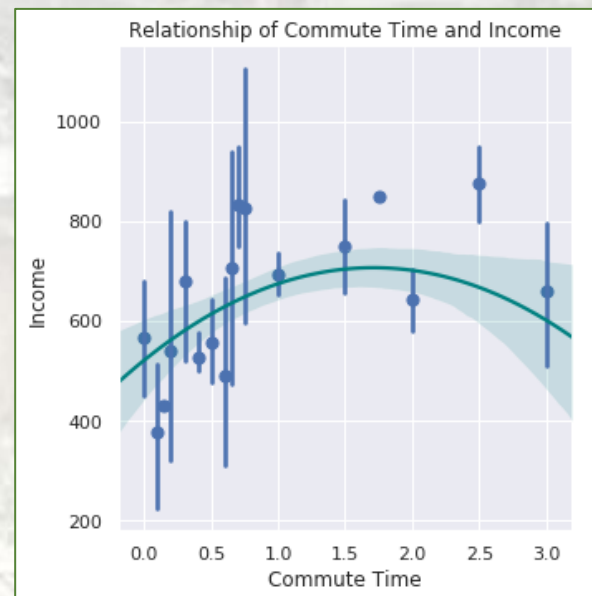
## Drill Down: Grandparents, Children and Commuters

Across various analyses, we began to notice an interconnected pattern among our insights which provided for a “voilà” moment. The idea came from noticing the strength of the primary earners in each household, the relationship between income and commute time, as well as a connectedness between income and presence of elderly in the household. Now, you ask, “How are these connected?” Well, let’s take it step by step:



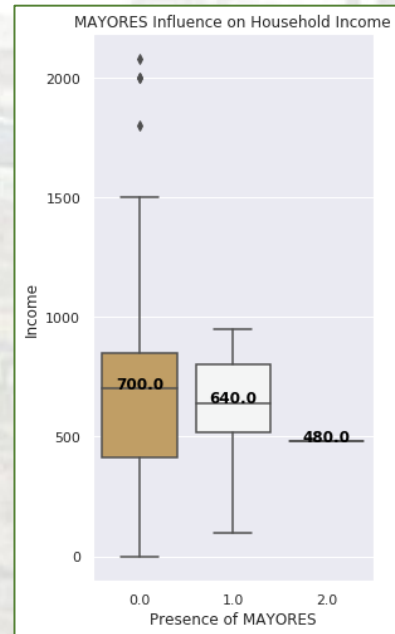
The number of workers in each household had low/positive coefficients and provided little predictive power in the determining of the income in the household. Thus, we produced a boxplot to see exactly how important the second and third earners are in each household. The results affirmed our previous analysis, insofar as additional workers in the household only provide a marginal increase in top-line household income, while the first/top earner in each household provides a majority of the income.

The second relationship of interest is that of commute time and income. The ways in which these variables interact offers a glimpse into the distance the good jobs are in relation to the slums of Pamplona Alta. The graph suggests that the best jobs are 35 minutes to 2.5 hours away from their homes; and though there are many jobs nearby, none of them pay very well.





Next, the coefficients of our initial analyses suggests that the elderly apply downward pressure on household income, however, a more in depth analysis discovered that the elderly actually serve as a normalizing factor to income. The average income of households with elderly are a little lower, yet the minimum income is actually much higher when they are present. This affirms what we know intrinsically about households in Perú, that multi-generational households are more of a source of insurance, rather than liability.



**Child Coefficient on Income**  
-16.13

Lastly, as anticipated, we see that the more kids someone has, the more downward pressure is applied to household income.

All things considered, we began to see the possibility for a marketplace within Pamplona Alta that could provide additional income to all households involved.

**Let's explain:** Most households have one primary earner, and to earn a better salary that person has to travel a substantial distance to get to that job. The more children in the household, the lower the top-line income of the family, however, the presence of older adults provides stability to family income.

**Market Opportunity:** We see an opportunity to create a marketplace inside communities for primary earners (especially single mothers) to accept jobs deeper into Lima that offer better income, after which, they can use a small portion of that increase in income to utilize elderly and underemployed in the local community to provide childcare.

**Financial Opportunity:** There is potential for a substantial increase household income for both the family with increased freedom to attain a better job deeper in Lima, as well as for the family who has the chance for supplemental income by providing childcare. An implementation challenge would be that a system for such a marketplace in Pamplona Alta would likely have to be SMS based, due to the low rates of phones with internet (11.06%).

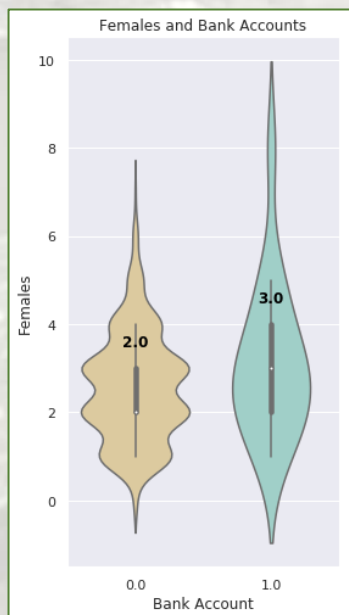
## Drill Down: Females and Finances

Internet and bank accounts are significant indicators of increased wealth and stability. Access to these resources can help create a digital footprint for families who are otherwise living in an invisible space and operating within a dark economy. Our analysis suggests that the wealthiest of those in our survey are the ones who utilize internet and bank services, though these services did not specifically cause the wealth.

Access to bank and internet services can give families increased visibility for loans, greater opportunity for savings and additional tools of validation in their fight for property ownership in Pamplona Alta. These tools can allow for better participation in the modern world economy and increase the potential tools available to better their lives

We performed a logistic regression to predict if someone had a bank account, and once again, the gold was in the coefficients. The more females in the house, the more likely the family is to have a bank account. The younger the mother (at the time of her first child) and the more children, the less likely a family is to have a bank account; implying that it may not just be financial scarcity problem, but a time scarcity issue that keeps families from getting a bank account.

F	1.063943
NINOS	-1.052686
primer_hijo	-0.965154
M	0.763479
ingreso	0.717030
tiempo_casa	-0.569880
PADRE	-0.269541
iglesia	-0.242704
dejar_hijos	0.210191
MADRE	-0.177087



In order to add some visual clarity to the impact of females and bank accounts we created a violin plot which shows us that households with bank accounts have more females in them. As we look to solution ideas such as financial classes, we will likely look to women first, because they are more likely to utilize these accounts.



## Drill Down: Newly Established Households

New members to the community struggle to make it to church, to get baptized and to gain access water more than established community members. In order to understand the reality of these newer households, multiple types of analyses are needed to give shape to this reality within the complex socio-political environment of Pamplona Alta.

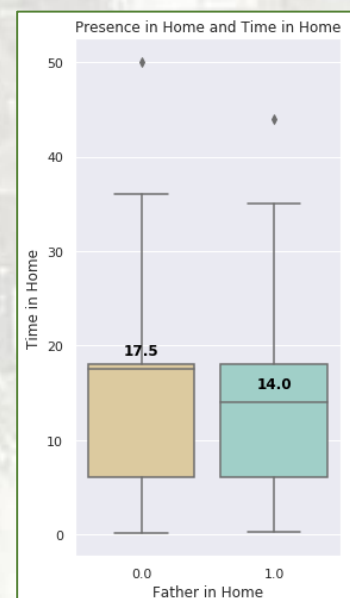
The coefficients of time in home resulting from the multiple linear regression results tell an interesting story. The longer someone has been in their home, the more likely they are to have water access, be baptized and attend church, but also the less likely they are to have fathers at home. For new households, work is also further from home than the more established community members.

	coef
internet	-0.9247
agua	1.6294
banco	-1.9254
iglesia	1.3106
dejar_hijos	-0.4571
cuantas_personas	1.9891
primer_hijo	0.1809
cuantas_trabajan	0.3203
tiempo_trabajan	-0.8203
pierden_colegio	-0.1750
ingreso	-0.0005
bautizadas	6.6100
PADRE	-3.0789
MADRE	-0.7539
F	0.2169
M	0.7064
NINOS	-1.5300
MAYORES	-2.9270



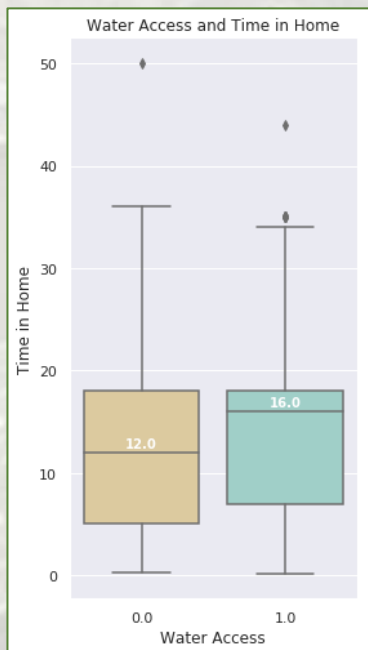
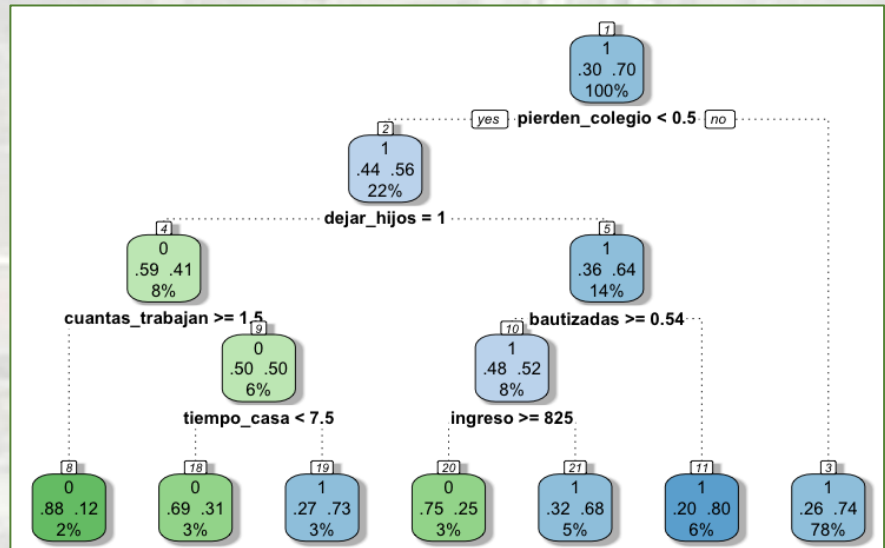
The complication surrounding the state of new households can seen in the relationship between income and time in home. There is essentially no relationship between the two, yet the coefficients of the multiple regression show that there are in fact many different upward and downward forces acting upon the income of new households all at once.

Multiple analyses have revealed that newly established homes are more likely to have a father at home and this boxplot allows us to better visualize this effect. The longer someone is in the household, the more likely the father of the household will leave. This makes intrinsic sense, however, the trick is to see what we do from here. We have proven the value of the father to income and stability; the goal is then be to create resources for young/new families to keep the father in the home.



Beyond income and family factors, we performed multiple analyses to understand the impact of water access on families and the subsequent results continually pointed to one thing, time in home. The most revealing of these analyses was the prediction of water access with a decision tree algorithm. A decision tree is a flow structure where each internal node represents a test on an factor and each branch is the outcome of each test. Additionally, every leaf node represents the final classification result. The path from the root to the leaf is the rules to that specific classification.

Looking at the result of the water access classification decision tree we can see that the story of water access is not as straightforward as we had anticipated. Those who do not have water access are those who have more workers at home, leave the house more and have been in their home for less than seven years. Given everything we know thus far, these are all characteristics of newly established households.



Coupled with our experience working in Pamplona Alta, we know that new members traditionally live higher on the mountain, and the water truck cannot access these homes. In order to isolate water access and time in home, we created a boxplot which shows that those with water access have been in their homes an average of 4 years longer than those who do not; affirming our beliefs.



## Conclusion: Summarized Solutions

We have worked hard to give shape to the community of Pamplona Alta, the problems they face and the specificities of their reality. Step by step, we have highlighted problems and implied a few possible solutions, however, many of these solutions deserve further elaboration.

**Cluster.** Our analysis of the types of people that exist within Pamplona Alta show that there are 4 different realities. An action step that can be taken is to outline solutions grouped by cluster in order to better help the specific groups as they need it, rather than applying the same kind of help to different groups.

**Communities.** In an area with unique (un)involvement of government entities, the community structure is integral to the health of the local area. Statistics and problem sets assessed on the community level allow us to see which communities need the most help, and in which areas.

**Household Size and Income.** Every household has a primary earner that carries an average family size of 5. We need to create resources and tools available to help stabilize the family financial structure, such as adding resources that increase financial stability.

**Role of Men.** The presence of men in the home provides stability and income that can support larger families. Men's groups and faith services could provide a foundation to keep men in the home, especially fathers of newly established homes.

**Grandparents, Children and Commuters.** The best jobs are an 35mins to 2.5hrs from the Pamplona Alta, and the further the job, the more likely the need for child care support. In addition to a high number of under/un employed adults at home, an SMS marketplace could exist to watch children, employ locals and increase earning power of parents.

**Females and Finance.** The more women in the household, the more likely a family is to have a bank account. The younger the family and the more kids, the less likely someone is to have a bank account; highlighting a scarcity of time rather than money in the attainment of a bank account. Financial and internet classes aimed at women could be available to service this need.

**Newly Established Households.** Members who have been in their home a shorter amount of time are more likely to have fathers at home, less likely to have water access, less likely to be involved in church services, yet not particularly poorer than other community members. These families could use church support, men could use support and paths to water access could be established.

The goal of this document is to offer problems with some solutions, however, the description of the problem is the primary result of this document. We hope that different organizations, NGOs and church entities can utilize this document and the resulting datasets to better help those in Pamplona Alta. Thank you to all those who helped in its creation.