# CSP 571: Project – Final Report

# Banking Product Recommendation

Gyanesh Tiwari(A20481772)| Shwejan Peddi(A20520861)|Venkat Gopi Yamparala(A20521252)|Yeswanth Sai Vuddula(A20519889)

## Abstract

**Personalized product recommendations are used by banks to reduce customer churn and maximize profit. We built one such personalized product recommendation system for Santander Bank using machine learning techniques and historical data. Initially, we planned to build four different machine learning modes - Naive Bayes, random forest, XGBoost and collaborative filtering, but we could deliver just two models – xgboost: multiclass and xgboost: binary. We have compared the performance of these two models using MAP@7 metric.**

### 1.1 Overview

Since it costs significantly more to acquire new customers than it does to keep existing ones, banks proactively offer new offers and individualized services to current clients. A sample of such a service is customized product recommendations. Through tailored product recommendations, Santander Bank extends a helping hand to its clients. A few of Santander's clients benefit greatly from their present approach, but many others receive few recommendations, creating an unequal customer experience. Santander can better satisfy the unique demands of every customer and guarantee their pleasure regardless of where they are in life with the implementation of an effective recommendation system.

### 1.2 Objective

The objective of this project is to create a machine learning-based solution leveraging past customer data to give Santander Bank with individualized product recommendations. It also enables us to answer the following intriguing queries regarding user behavior.
We seek to answer the following questions:

### Questions addressed by the Project

- Which products are most popular, and which are least popular?

- How does the popularity for a product change over time?
- Demographic information about the customer?

**2. Data Processing**

Our dataset is from Kaggle competition for Santander bank. It contains both train and test files. Train dataset contains up to 17 months of data for each customer and has 13647309 records with 48 columns. Test dataset contains 1 month of data and has 929615 records with 48 columns. First 24 columns represent features and next 24 features represent products. We club both train and test together to perform EDA. Out of these 24 features, there is one id column(ncodpers), three date columns (fecha_dato, fecha_alta, ult_fec_cli_lt), 3 numeric columns(age, antiguedad, renta), 17 categorical columns( ind_empleado, pais_residencia, sexo, ind_nuevo, indrel, indrel_1mes, tiprel_1mes, indresi, indext, conyuemp, canal_entrada, indfall, tipodom, cod_prov, nomprov, ind_actividad_cliente, segmento). There are many inconsistencies in data such as missing values, formatting issues. We have addresses each of these below.

DataSet is available at below location:
https://www.kaggle.com/competitions/santander-product-recommendation/data

**2.1 Date and Time Column**

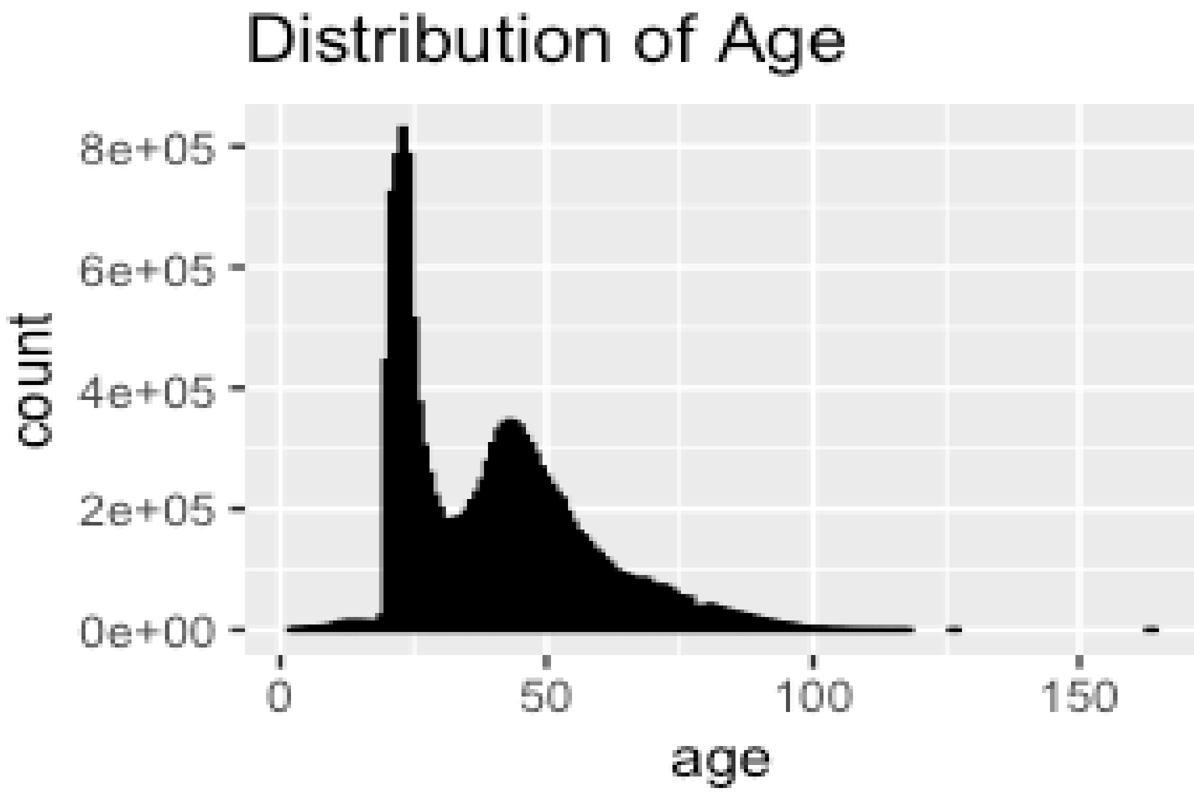
There are three date columns – fecha_dato, fecha_alta, ult_fec_cli_lt. We converted these columns into YYYY-DD-MM.

**2.2 Missing value analysis and Imputation**

There are 24 columns where values are missing. To impute missing values, we have followed following approach:

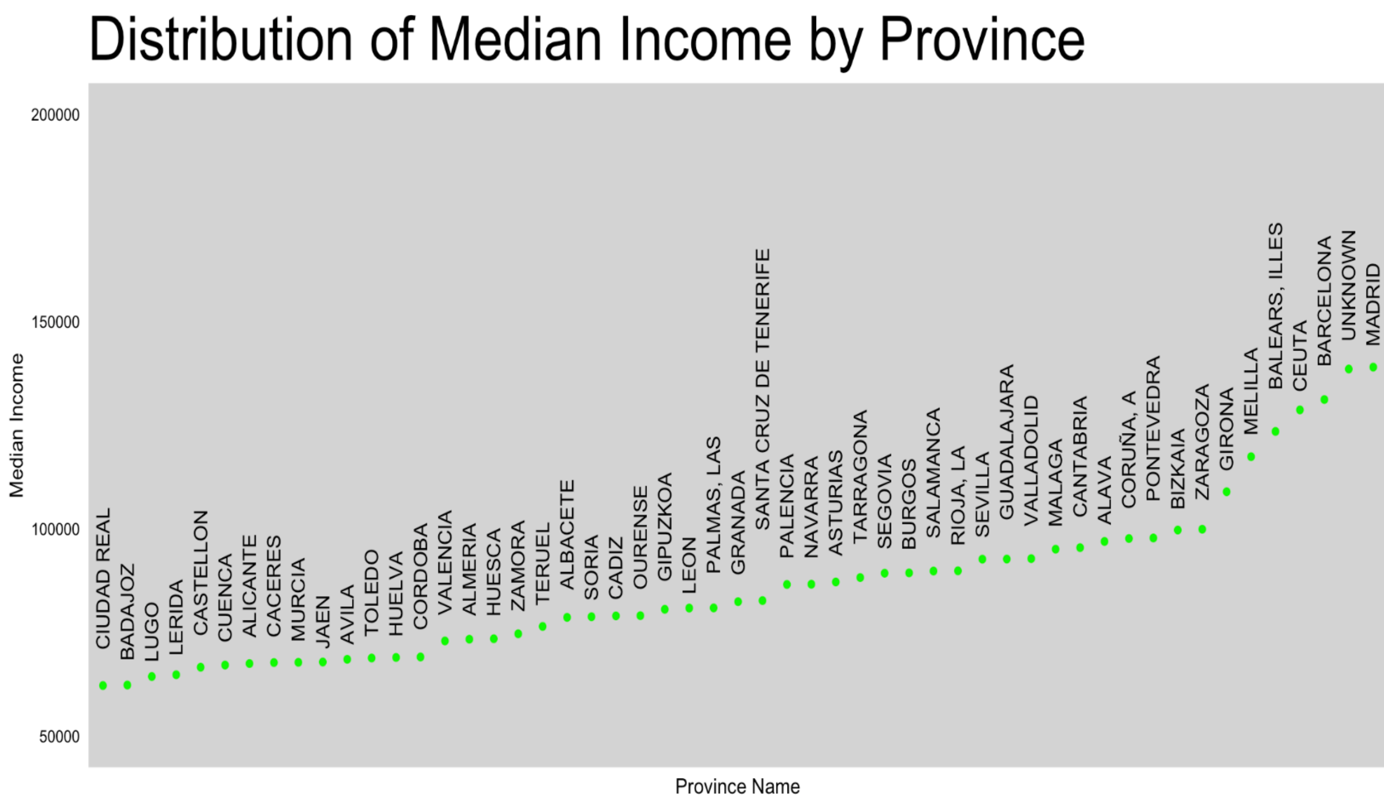
- For numeric columns, we observe the distribution and find a suitable value
- For categorical columns, impute with majority class if there is only one majority class else impute with a new category “UNKNOWN”.
- For date columns, impute with median date

Let’s explore each of these columns:
**ind_empleado**: Impute missing values with majority class ‘N’.
**pais_residencia**: Impute missing values with majority class ‘ES’.
**sexo**: Impute missing values with majority class ‘V’.
**age:** Minimum value of age is 2 and maximum value is 164. It is bimodal in nature with first peak at 23 and second at 43. We imputed missing age with median value.




**ind_nuevo:** Since maximum number of records for a customer with missing ind_nuevo is 6 hence impute with 1.
**antiguedad:** Since these are the same customers which had missing for ind_neuvo values. Missing values are given minimum seniority.
**fecha_alta:** Impute missing date with median date of the column.
**Indrel:** Impute missing value with majority value 1.
**ult_fec_cli_1t:** Drop this column as 99.8 percent of the values are missing.
**indrel_1mes:** Impute with majority class "1".
**tiprel_1mes:** Impute with majority category "I".
**indresi:** Impute with majority category "S".
**indext**: Impute with majority category "N".
**conyuemp:** Impute with a new category "UNKNOWN".
**canal_entrada:** Impute with a new category "UNKNOWN".
**indfall:** Impute with majority class "N".
**tipodom:** Drop this column as it has only one value (zero variance).
**cod_prov:** Drop this column as next column is province name which is more descriptive.
**nomprov:** Impute with a new category "UNKNOWN".
**ind_actividad_cliente:** Impute with majority category 0.
**renta:** It has a minimum value of 1202.73 and a maximum value of 28894396. Since this range is

very high, we first observe the median renta for each province. There is a lot of variation in median income for different provinces. Hence it is imputed by median renta for each province.




**ind_nomina_ult1:** Impute missing value with majority class 0.
**Segmento:** Impute missing values with new category “UNKNOWN”.
**ind_nom_pens_ult1:** Impute with the majority class 0.

**2.3 New Feature Creation**

**month_id:** This is numeric feature which is derived from fecha_dato and has values 1 to 18.
**previous_month_id:** This is a numeric feature which equals month_id – 1.
**birthday_month:** This is a numeric feature. Customers tend to buy new products near their birthday.
**month:** This is a categorical feature and represents month of the transaction.
**product:** This column contains all the products that were bought in that month. If a product was present in the previous month and continued in the current month it is call maintained (which is not a concern for us). We are just focusing on products which were not present in the previous month but present in current month.

**activity_index_change:** Flag with value 0 if ind_actividad_cliente is same for current and previous month, else 1
**segmento_change:** Flag with value 0 if segmento is same for current and previous month, else 1

**Lag Feature Creation**

Lag features refer to lagged product ownership (Whether or not the product was owned 1,2,3,4,5 etc. months ago). For each product, it is beneficial to consider not only it's value for the current month but also the value for previous months.

**lagged_ind_actividad_cliente.1months_ago:** 1 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.2months_ago:** 2 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.3months_ago:** 3 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.4months_ago:** 4 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.5months_ago:** 5 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.6months_ago:** 6 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.7months_ago:** 7 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.8months_ago:** 8 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.9months_ago:** 9 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.10months_ago:** 10 month lag for ind_actividad_cliente
**lagged_ind_actividad_cliente.11months_ago:** 11 month lag for ind_actividad_cliente

**num_purchases_1_months_ago:** number of new products purchased 1 months ago
**num_purchases_2_months_ago:** number of new products purchased 2 months ago
**num_purchases_3_months_ago:** number of new products purchased 3 months ago
**num_purchases_4_months_ago:** number of new products purchased 4 months ago

**total_products_1_months_ago:** total products owned 1 months ago
**total_products_2_months_ago:** total products owned 2 months ago
**total_products_3_months_ago:** total products owned 3 months ago
**total_products_4_months_ago:** total products owned 4 months ago
**total_products_5_months_ago:** total products owned 5 months ago

**ind_ahor_fin_ult1_purchase_count:** Number of times product ind_ahor_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_aval_fin_ult1_purchase_count:** Number of times product ind_aval_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_cco_fin_ult1_purchase_count:** Number of times product ind_cco_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_cder_fin_ult1_purchase_count:** Number of times product ind_cder_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_cno_fin_ult1_purchase_count:** Number of times product ind_cno_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_ctju_fin_ult1_purchase_count:** Number of times product ind_ctju_fin_ult1 has been

purchases by a customer since Jan 2015.
**ind_ctma_fin_ult1_purchase_count:** Number of times product ind_ctma_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_ctop_fin_ult1_purchase_count:** Number of times product ind_ctop_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_ctpp_fin_ult1_purchase_count:** Number of times product ind_ctpp_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_deco_fin_ult1_purchase_count:** Number of times product ind_deco_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_deme_fin_ult1_purchase_count:** Number of times product ind_deme_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_dela_fin_ult1_purchase_count:** Number of times product ind_dela_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_ecue_fin_ult1_purchase_count:** Number of times product ind_ecue_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_fond_fin_ult1_purchase_count:** Number of times product ind_fond_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_hip_fin_ult1_purchase_count:** Number of times product ind_hip_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_plan_fin_ult1_purchase_count:** Number of times product ind_plan_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_pres_fin_ult1_purchase_count:** Number of times product ind_pres_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_reca_fin_ult1_purchase_count:** Number of times product ind_reca_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_tjcr_fin_ult1_purchase_count:** Number of times product ind_tjcr_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_valo_fin_ult1_purchase_count:** Number of times product ind_valo_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_viv_fin_ult1_purchase_count:** Number of times product ind_viv_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_nomina_ult1_purchase_count:** Number of times product ind_nomina_fin_ult1 has been purchases by a customer since Jan 2015.
**ind_nom_pens_ult1_purchase_count:** Number of times product ind_nom_pens_ult1 has been purchases by a customer since Jan 2015.
**ind_recibo_ult1_purchase_count:** Number of times product ind_recibo_ult1 has been purchases by a customer since Jan 2015.
**num_transactions:** Number of transactions each month.

Product ownership 1 to 11 months ago: For each of the 24 products we have captured it's lag from 1 to 11 months. Below is the name of those features.

**ind_ahor_fin_ult1_1month_ago:** ind_ahor_fin_ult1 ownership 1 month ago.
**ind_aval_fin_ult1_1month_ago:** ind_aval_fin_ult1 ownership 1 month ago.

**ind_cco_fin_ult1_1month_ago:** ind_cco_fin_ult1 ownership 1 month ago.
**ind_cder_fin_ult1_1month_ago:** ind_cder_fin_ult1 ownership 1 month ago.
**ind_cno_fin_ult1_1month_ago:** ind_cno_fin_ult1 ownership 1 month ago.
**ind_ctju_fin_ult1_1month_ago:** ind_ctju_fin_ult1 ownership 1 month ago.
**ind_ctma_fin_ult1_1month_ago:** ind_ctma_fin_ult1 ownership 1 month ago.
**ind_ctop_fin_ult1_1month_ago:** ind_ctop_fin_ult1 ownership 1 month ago.
**ind_ctpp_fin_ult1_1month_ago:** ind_ctpp_fin_ult1 ownership 1 month ago.
**ind_deco_fin_ult1_1month_ago:** ind_deco_fin_ult1 ownership 1 month ago.
**ind_deme_fin_ult1_1month_ago:** ind_deme_fin_ult1 ownership 1 month ago.
**ind_dela_fin_ult1_1month_ago:** ind_dela_fin_ult1 ownership 1 month ago.
**ind_ecue_fin_ult1_1month_ago:** ind_ecue_fin_ult1 ownership 1 month ago.
**ind_fond_fin_ult1_1month_ago:** ind_fond_fin_ult1 ownership 1 month ago.
**ind_hip_fin_ult1_1month_ago:** ind_hip_fin_ult1 ownership 1 month ago.
**ind_plan_fin_ult1_1month_ago:** ind_plan_fin_ult1 ownership 1 month ago.
**ind_pres_fin_ult1_1month_ago:** ind_pres_fin_ult1 ownership 1 month ago.
**ind_reca_fin_ult1_1month_ago:** ind_reca_fin_ult1 ownership 1 month ago.
**ind_tjcr_fin_ult1_1month_ago:** ind_tjcr_fin_ult1 ownership 1 month ago.
**ind_valo_fin_ult1_1month_ago:** ind_valo_fin_ult1 ownership 1 month ago.
**ind_viv_fin_ult1_1month_ago:** ind_viv_fin_ult1 ownership 1 month ago.
**ind_nomina_ult1_1month_ago:** ind_nomina_ult1 ownership 1 month ago.
**ind_nom_pens_ult1_1month_ago:** ind_nom_pens_ult1 ownership 1 month ago.
**ind_recibo_ult1_1month_ago:** ind_recibo_ult1 ownership 1 month ago.

**ind_ahor_fin_ult1_2month_ago:** ind_ahor_fin_ult1 ownership 2 months ago.
**ind_aval_fin_ult1_2month_ago:** ind_aval_fin_ult1 ownership 2 months ago.
**ind_cco_fin_ult1_2month_ago:** ind_cco_fin_ult1 ownership 2 months ago.
**ind_cder_fin_ult1_2month_ago:** ind_cder_fin_ult1 ownership 2 months ago.
**ind_cno_fin_ult1_2month_ago:** ind_cno_fin_ult1 ownership 2 months ago.
**ind_ctju_fin_ult1_2month_ago:** ind_ctju_fin_ult1 ownership 2 months ago.
**ind_ctma_fin_ult1_2month_ago:** ind_ctma_fin_ult1 ownership 2 months ago.
**ind_ctop_fin_ult1_2month_ago:** ind_ctop_fin_ult1 ownership 2 months ago.
**ind_ctpp_fin_ult1_2month_ago:** ind_ctpp_fin_ult1 ownership 2 months ago.
**ind_deco_fin_ult1_2month_ago:** ind_deco_fin_ult1 ownership 2 months ago.
**ind_deme_fin_ult1_2month_ago:** ind_deme_fin_ult1 ownership 2 months ago.
**ind_dela_fin_ult1_2month_ago:** ind_dela_fin_ult1 ownership 2 months ago.
**ind_ecue_fin_ult1_2month_ago:** ind_ecue_fin_ult1 ownership 2 months ago.
**ind_fond_fin_ult1_2month_ago:** ind_fond_fin_ult1 ownership 2 months ago.
**ind_hip_fin_ult1_2month_ago:** ind_hip_fin_ult1 ownership 2 months ago.
**ind_plan_fin_ult1_2month_ago:** ind_plan_fin_ult1 ownership 2 months ago.
**ind_pres_fin_ult1_2month_ago:** ind_pres_fin_ult1 ownership 2 months ago.
**ind_reca_fin_ult1_2month_ago:** ind_reca_fin_ult1 ownership 2 months ago.
**ind_tjcr_fin_ult1_2month_ago:** ind_tjcr_fin_ult1 ownership 2 months ago.
**ind_valo_fin_ult1_2month_ago:** ind_valo_fin_ult1 ownership 2 months ago.
**ind_viv_fin_ult1_2month_ago:** ind_viv_fin_ult1 ownership 2 months ago.

**ind_nomina_ult1_2month_ago:** ind_nomina_ult1 ownership 2 months ago.
**ind_nom_pens_ult1_2month_ago:** ind_nom_pens_ult1 ownership 2 months ago.
**ind_recibo_ult1_2month_ago:** ind_recibo_ult1 ownership 2 months ago.

**ind_ahor_fin_ult1_3month_ago:** ind_ahor_fin_ult1 ownership 3 months ago.
**ind_aval_fin_ult1_3month_ago:** ind_aval_fin_ult1 ownership 3 months ago.
**ind_cco_fin_ult1_3month_ago:** ind_cco_fin_ult1 ownership 3 months ago.
**ind_cder_fin_ult1_3month_ago:** ind_cder_fin_ult1 ownership 3 months ago.
**ind_cno_fin_ult1_3month_ago:** ind_cno_fin_ult1 ownership 3 months ago.
**ind_ctju_fin_ult1_3month_ago:** ind_ctju_fin_ult1 ownership 3 months ago.
**ind_ctma_fin_ult1_3month_ago:** ind_ctma_fin_ult1 ownership 3 months ago.
**ind_ctop_fin_ult1_3month_ago:** ind_ctop_fin_ult1 ownership 3 months ago.
**ind_ctpp_fin_ult1_3month_ago:** ind_ctpp_fin_ult1 ownership 3 months ago.
**ind_deco_fin_ult1_3month_ago:** ind_deco_fin_ult1 ownership 3 months ago.
**ind_deme_fin_ult1_3month_ago:** ind_deme_fin_ult1 ownership 3 months ago.
**ind_dela_fin_ult1_3month_ago:** ind_dela_fin_ult1 ownership 3 months ago.
**ind_ecue_fin_ult1_3month_ago:** ind_ecue_fin_ult1 ownership 3 months ago.
**ind_fond_fin_ult1_3month_ago:** ind_fond_fin_ult1 ownership 3 months ago.
**ind_hip_fin_ult1_3month_ago:** ind_hip_fin_ult1 ownership 3 months ago.
**ind_plan_fin_ult1_3month_ago:** ind_plan_fin_ult1 ownership 3 months ago.
**ind_pres_fin_ult1_3month_ago:** ind_pres_fin_ult1 ownership 3 months ago.
**ind_reca_fin_ult1_3month_ago:** ind_reca_fin_ult1 ownership 3 months ago.
**ind_tjcr_fin_ult1_3month_ago:** ind_tjcr_fin_ult1 ownership 3 months ago.
**ind_valo_fin_ult1_3month_ago:** ind_valo_fin_ult1 ownership 3 months ago.
**ind_viv_fin_ult1_3month_ago:** ind_viv_fin_ult1 ownership 3 months ago.
**ind_nomina_ult1_3month_ago:** ind_nomina_ult1 ownership 3 months ago.
**ind_nom_pens_ult1_3month_ago:** ind_nom_pens_ult1 ownership 3 months ago.
**ind_recibo_ult1_3month_ago:** ind_recibo_ult1 ownership 3 months ago.

**ind_ahor_fin_ult1_4month_ago:** ind_ahor_fin_ult1 ownership 4 months ago.
**ind_aval_fin_ult1_4month_ago:** ind_aval_fin_ult1 ownership 4 months ago.
**ind_cco_fin_ult1_4month_ago:** ind_cco_fin_ult1 ownership 4 months ago.
**ind_cder_fin_ult1_4month_ago:** ind_cder_fin_ult1 ownership 4 months ago.
**ind_cno_fin_ult1_4month_ago:** ind_cno_fin_ult1 ownership 4 months ago.
**ind_ctju_fin_ult1_4month_ago:** ind_ctju_fin_ult1 ownership 4 months ago.
**ind_ctma_fin_ult1_4month_ago:** ind_ctma_fin_ult1 ownership 4 months ago.
**ind_ctop_fin_ult1_4month_ago:** ind_ctop_fin_ult1 ownership 4 months ago.
**ind_ctpp_fin_ult1_4month_ago:** ind_ctpp_fin_ult1 ownership 4 months ago.
**ind_deco_fin_ult1_4month_ago:** ind_deco_fin_ult1 ownership 4 months ago.
**ind_deme_fin_ult1_4month_ago:** ind_deme_fin_ult1 ownership 4 months ago.
**ind_dela_fin_ult1_4month_ago:** ind_dela_fin_ult1 ownership 4 months ago.
**ind_ecue_fin_ult1_4month_ago:** ind_ecue_fin_ult1 ownership 4 months ago.

**ind_fond_fin_ult1_4month_ago:** ind_fond_fin_ult1 ownership 4 months ago.
**ind_hip_fin_ult1_4month_ago:** ind_hip_fin_ult1 ownership 4 months ago.
**ind_plan_fin_ult1_4month_ago:** ind_plan_fin_ult1 ownership 4 months ago.
**ind_pres_fin_ult1_4month_ago:** ind_pres_fin_ult1 ownership 4 months ago.
**ind_reca_fin_ult1_4month_ago:** ind_reca_fin_ult1 ownership 4 months ago.
**ind_tjcr_fin_ult1_4month_ago:** ind_tjcr_fin_ult1 ownership 4 months ago.
**ind_valo_fin_ult1_4month_ago:** ind_valo_fin_ult1 ownership 4 months ago.
**ind_viv_fin_ult1_4month_ago:** ind_viv_fin_ult1 ownership 4 months ago.
**ind_nomina_ult1_4month_ago:** ind_nomina_ult1 ownership 4 months ago.
**ind_nom_pens_ult1_4month_ago:** ind_nom_pens_ult1 ownership 4 months ago.
**ind_recibo_ult1_4month_ago:** ind_recibo_ult1 ownership 4 months ago.

**ind_ahor_fin_ult1_5month_ago:** ind_ahor_fin_ult1 ownership 5 months ago.
**ind_aval_fin_ult1_5month_ago:** ind_aval_fin_ult1 ownership 5 months ago.
**ind_cco_fin_ult1_5month_ago:** ind_cco_fin_ult1 ownership 5 months ago.
**ind_cder_fin_ult1_5month_ago:** ind_cder_fin_ult1 ownership 5 months ago.
**ind_cno_fin_ult1_5month_ago:** ind_cno_fin_ult1 ownership 5 months ago.
**ind_ctju_fin_ult1_5month_ago:** ind_ctju_fin_ult1 ownership 5 months ago.
**ind_ctma_fin_ult1_5month_ago:** ind_ctma_fin_ult1 ownership 5 months ago.
**ind_ctop_fin_ult1_5month_ago:** ind_ctop_fin_ult1 ownership 5 months ago.
**ind_ctpp_fin_ult1_5month_ago:** ind_ctpp_fin_ult1 ownership 5 months ago.
**ind_deco_fin_ult1_5month_ago:** ind_deco_fin_ult1 ownership 5 months ago.
**ind_deme_fin_ult1_5month_ago:** ind_deme_fin_ult1 ownership 5 months ago.
**ind_dela_fin_ult1_5month_ago:** ind_dela_fin_ult1 ownership 5 months ago.
**ind_ecue_fin_ult1_5month_ago:** ind_ecue_fin_ult1 ownership 5 months ago.
**ind_fond_fin_ult1_5month_ago:** ind_fond_fin_ult1 ownership 5 months ago.
**ind_hip_fin_ult1_5month_ago:** ind_hip_fin_ult1 ownership 5 months ago.
**ind_plan_fin_ult1_5month_ago:** ind_plan_fin_ult1 ownership 5 months ago.
**ind_pres_fin_ult1_5month_ago:** ind_pres_fin_ult1 ownership 5 months ago.
**ind_reca_fin_ult1_5month_ago:** ind_reca_fin_ult1 ownership 5 months ago.
**ind_tjcr_fin_ult1_5month_ago:** ind_tjcr_fin_ult1 ownership 5 months ago.
**ind_valo_fin_ult1_5month_ago:** ind_valo_fin_ult1 ownership 5 months ago.
**ind_viv_fin_ult1_5month_ago:** ind_viv_fin_ult1 ownership 5 months ago.
**ind_nomina_ult1_5month_ago:** ind_nomina_ult1 ownership 5 months ago.
**ind_nom_pens_ult1_5month_ago:** ind_nom_pens_ult1 ownership 5 months ago.
**ind_recibo_ult1_5month_ago:** ind_recibo_ult1 ownership 5 months ago.

**ind_ahor_fin_ult1_6month_ago:** ind_ahor_fin_ult1 ownership 6 months ago.
**ind_aval_fin_ult1_6month_ago:** ind_aval_fin_ult1 ownership 6 months ago.
**ind_cco_fin_ult1_6month_ago:** ind_cco_fin_ult1 ownership 6 months ago.
**ind_cder_fin_ult1_6month_ago:** ind_cder_fin_ult1 ownership 6 months ago.
**ind_cno_fin_ult1_6month_ago:** ind_cno_fin_ult1 ownership 6 months ago.

**ind_ctju_fin_ult1_6month_ago:** ind_ctju_fin_ult1 ownership 6 months ago.
**ind_ctma_fin_ult1_6month_ago:** ind_ctma_fin_ult1 ownership 6 months ago.
**ind_ctop_fin_ult1_6month_ago:** ind_ctop_fin_ult1 ownership 6 months ago.
**ind_ctpp_fin_ult1_6month_ago:** ind_ctpp_fin_ult1 ownership 6 months ago.
**ind_deco_fin_ult1_6month_ago:** ind_deco_fin_ult1 ownership 6 months ago.
**ind_deme_fin_ult1_6month_ago:** ind_deme_fin_ult1 ownership 6 months ago.
**ind_dela_fin_ult1_6month_ago:** ind_dela_fin_ult1 ownership 6 months ago.
**ind_ecue_fin_ult1_6month_ago:** ind_ecue_fin_ult1 ownership 6 months ago.
**ind_fond_fin_ult1_6month_ago:** ind_fond_fin_ult1 ownership 6 months ago.
**ind_hip_fin_ult1_6month_ago:** ind_hip_fin_ult1 ownership 6 months ago.
**ind_plan_fin_ult1_6month_ago:** ind_plan_fin_ult1 ownership 6 months ago.
**ind_pres_fin_ult1_6month_ago:** ind_pres_fin_ult1 ownership 6 months ago.
**ind_reca_fin_ult1_6month_ago:** ind_reca_fin_ult1 ownership 6 months ago.
**ind_tjcr_fin_ult1_6month_ago:** ind_tjcr_fin_ult1 ownership 6 months ago.
**ind_valo_fin_ult1_6month_ago:** ind_valo_fin_ult1 ownership 6 months ago.
**ind_viv_fin_ult1_6month_ago:** ind_viv_fin_ult1 ownership 6 months ago.
**ind_nomina_ult1_6month_ago:** ind_nomina_ult1 ownership 6 months ago.
**ind_nom_pens_ult1_6month_ago:** ind_nom_pens_ult1 ownership 6 months ago.
**ind_recibo_ult1_6month_ago:** ind_recibo_ult1 ownership 6 months ago.

**ind_ahor_fin_ult1_7month_ago:** ind_ahor_fin_ult1 ownership 7 months ago.
**ind_aval_fin_ult1_7month_ago:** ind_aval_fin_ult1 ownership 7 months ago.
**ind_cco_fin_ult1_7month_ago:** ind_cco_fin_ult1 ownership 7 months ago.
**ind_cder_fin_ult1_7month_ago:** ind_cder_fin_ult1 ownership 7 months ago.
**ind_cno_fin_ult1_7month_ago:** ind_cno_fin_ult1 ownership 7 months ago.
**ind_ctju_fin_ult1_7month_ago:** ind_ctju_fin_ult1 ownership 7 months ago.
**ind_ctma_fin_ult1_7month_ago:** ind_ctma_fin_ult1 ownership 7 months ago.
**ind_ctop_fin_ult1_7month_ago:** ind_ctop_fin_ult1 ownership 7 months ago.
**ind_ctpp_fin_ult1_7month_ago:** ind_ctpp_fin_ult1 ownership 7 months ago.
**ind_deco_fin_ult1_7month_ago:** ind_deco_fin_ult1 ownership 7 months ago.
**ind_deme_fin_ult1_7month_ago:** ind_deme_fin_ult1 ownership 7 months ago.
**ind_dela_fin_ult1_7month_ago:** ind_dela_fin_ult1 ownership 7 months ago.
**ind_ecue_fin_ult1_7month_ago:** ind_ecue_fin_ult1 ownership 7 months ago.
**ind_fond_fin_ult1_7month_ago:** ind_fond_fin_ult1 ownership 7 months ago.
**ind_hip_fin_ult1_7month_ago:** ind_hip_fin_ult1 ownership 7 months ago.
**ind_plan_fin_ult1_7month_ago:** ind_plan_fin_ult1 ownership 7 months ago.
**ind_pres_fin_ult1_7month_ago:** ind_pres_fin_ult1 ownership 7 months ago.
**ind_reca_fin_ult1_7month_ago:** ind_reca_fin_ult1 ownership 7 months ago.
**ind_tjcr_fin_ult1_7month_ago:** ind_tjcr_fin_ult1 ownership 7 months ago.
**ind_valo_fin_ult1_7month_ago:** ind_valo_fin_ult1 ownership 7 months ago.
**ind_viv_fin_ult1_7month_ago:** ind_viv_fin_ult1 ownership 7 months ago.
**ind_nomina_ult1_7month_ago:** ind_nomina_ult1 ownership 7 months ago.

**ind_nom_pens_ult1_7month_ago:** ind_nom_pens_ult1 ownership 7 months ago.
**ind_recibo_ult1_7month_ago:** ind_recibo_ult1 ownership 7 months ago.

**ind_ahor_fin_ult1_8month_ago:** ind_ahor_fin_ult1 ownership 8 months ago.
**ind_aval_fin_ult1_8month_ago:** ind_aval_fin_ult1 ownership 8 months ago.
**ind_cco_fin_ult1_8month_ago:** ind_cco_fin_ult1 ownership 8 months ago.
**ind_cder_fin_ult1_8month_ago:** ind_cder_fin_ult1 ownership 8 months ago.
**ind_cno_fin_ult1_8month_ago:** ind_cno_fin_ult1 ownership 8 months ago.
**ind_ctju_fin_ult1_8month_ago:** ind_ctju_fin_ult1 ownership 8 months ago.
**ind_ctma_fin_ult1_8month_ago:** ind_ctma_fin_ult1 ownership 8 months ago.
**ind_ctop_fin_ult1_8month_ago:** ind_ctop_fin_ult1 ownership 8 months ago.
**ind_ctpp_fin_ult1_8month_ago:** ind_ctpp_fin_ult1 ownership 8 months ago.
**ind_deco_fin_ult1_8month_ago:** ind_deco_fin_ult1 ownership 8 months ago.
**ind_deme_fin_ult1_8month_ago:** ind_deme_fin_ult1 ownership 8 months ago.
**ind_dela_fin_ult1_8month_ago:** ind_dela_fin_ult1 ownership 8 months ago.
**ind_ecue_fin_ult1_8month_ago:** ind_ecue_fin_ult1 ownership 8 months ago.
**ind_fond_fin_ult1_8month_ago:** ind_fond_fin_ult1 ownership 8 months ago.
**ind_hip_fin_ult1_8month_ago:** ind_hip_fin_ult1 ownership 8 months ago.
**ind_plan_fin_ult1_8month_ago:** ind_plan_fin_ult1 ownership 8 months ago.
**ind_pres_fin_ult1_8month_ago:** ind_pres_fin_ult1 ownership 8 months ago.
**ind_reca_fin_ult1_8month_ago:** ind_reca_fin_ult1 ownership 8 months ago.
**ind_tjcr_fin_ult1_8month_ago:** ind_tjcr_fin_ult1 ownership 8 months ago.
**ind_valo_fin_ult1_8month_ago:** ind_valo_fin_ult1 ownership 8 months ago.
**ind_viv_fin_ult1_8month_ago:** ind_viv_fin_ult1 ownership 8 months ago.
**ind_nomina_ult1_8month_ago:** ind_nomina_ult1 ownership 8 months ago.
**ind_nom_pens_ult1_8month_ago:** ind_nom_pens_ult1 ownership 8 months ago.
**ind_recibo_ult1_8month_ago:** ind_recibo_ult1 ownership 8 months ago.

**ind_ahor_fin_ult1_9month_ago:** ind_ahor_fin_ult1 ownership 9 months ago.
**ind_aval_fin_ult1_9month_ago:** ind_aval_fin_ult1 ownership 9 months ago.
**ind_cco_fin_ult1_9month_ago:** ind_cco_fin_ult1 ownership 9 months ago.
**ind_cder_fin_ult1_9month_ago:** ind_cder_fin_ult1 ownership 9 months ago.
**ind_cno_fin_ult1_9month_ago:** ind_cno_fin_ult1 ownership 9 months ago.
**ind_ctju_fin_ult1_9month_ago:** ind_ctju_fin_ult1 ownership 9 months ago.
**ind_ctma_fin_ult1_9month_ago:** ind_ctma_fin_ult1 ownership 9 months ago.
**ind_ctop_fin_ult1_9month_ago:** ind_ctop_fin_ult1 ownership 9 months ago.
**ind_ctpp_fin_ult1_9month_ago:** ind_ctpp_fin_ult1 ownership 9 months ago.
**ind_deco_fin_ult1_9month_ago:** ind_deco_fin_ult1 ownership 9 months ago.
**ind_deme_fin_ult1_9month_ago:** ind_deme_fin_ult1 ownership 9 months ago.
**ind_dela_fin_ult1_9month_ago:** ind_dela_fin_ult1 ownership 9 months ago.

**ind_ecue_fin_ult1_9month_ago:** ind_ecue_fin_ult1 ownership 9 months ago.
**ind_fond_fin_ult1_9month_ago:** ind_fond_fin_ult1 ownership 9 months ago.
**ind_hip_fin_ult1_9month_ago:** ind_hip_fin_ult1 ownership 9 months ago.
**ind_plan_fin_ult1_9month_ago:** ind_plan_fin_ult1 ownership 9 months ago.
**ind_pres_fin_ult1_9month_ago:** ind_pres_fin_ult1 ownership 9 months ago.
**ind_reca_fin_ult1_9month_ago:** ind_reca_fin_ult1 ownership 9 months ago.
**ind_tjcr_fin_ult1_9month_ago:** ind_tjcr_fin_ult1 ownership 9 months ago.
**ind_valo_fin_ult1_9month_ago:** ind_valo_fin_ult1 ownership 9 months ago.
**ind_viv_fin_ult1_9month_ago:** ind_viv_fin_ult1 ownership 9 months ago.
**ind_nomina_ult1_9month_ago:** ind_nomina_ult1 ownership 9 months ago.
**ind_nom_pens_ult1_9month_ago:** ind_nom_pens_ult1 ownership 9 months ago.
**ind_recibo_ult1_9month_ago:** ind_recibo_ult1 ownership 9 months ago.

**ind_ahor_fin_ult1_10month_ago:** ind_ahor_fin_ult1 ownership 10 months ago.
**ind_aval_fin_ult1_10month_ago:** ind_aval_fin_ult1 ownership 10 months ago.
**ind_cco_fin_ult1_10month_ago:** ind_cco_fin_ult1 ownership 10 months ago.
**ind_cder_fin_ult1_10month_ago:** ind_cder_fin_ult1 ownership 10 months ago.
**ind_cno_fin_ult1_10month_ago:** ind_cno_fin_ult1 ownership 10 months ago.
**ind_ctju_fin_ult1_10month_ago:** ind_ctju_fin_ult1 ownership 10 months ago.
**ind_ctma_fin_ult1_10month_ago:** ind_ctma_fin_ult1 ownership 10 months ago.
**ind_ctop_fin_ult1_10month_ago:** ind_ctop_fin_ult1 ownership 10 months ago.
**ind_ctpp_fin_ult1_10month_ago:** ind_ctpp_fin_ult1 ownership 10 months ago.
**ind_deco_fin_ult1_10month_ago:** ind_deco_fin_ult1 ownership 10 months ago.
**ind_deme_fin_ult1_10month_ago:** ind_deme_fin_ult1 ownership 10 months ago.
**ind_dela_fin_ult1_10month_ago:** ind_dela_fin_ult1 ownership 10 months ago.
**ind_ecue_fin_ult1_10month_ago:** ind_ecue_fin_ult1 ownership 10 months ago.
**ind_fond_fin_ult1_10month_ago:** ind_fond_fin_ult1 ownership 10 months ago.
**ind_hip_fin_ult1_10month_ago:** ind_hip_fin_ult1 ownership 10 months ago.
**ind_plan_fin_ult1_10month_ago:** ind_plan_fin_ult1 ownership 10 months ago.
**ind_pres_fin_ult1_10month_ago:** ind_pres_fin_ult1 ownership 10 months ago.
**ind_reca_fin_ult1_10month_ago:** ind_reca_fin_ult1 ownership 10 months ago.
**ind_tjcr_fin_ult1_10month_ago:** ind_tjcr_fin_ult1 ownership 10 months ago.
**ind_valo_fin_ult1_10month_ago:** ind_valo_fin_ult1 ownership 10 months ago.
**ind_viv_fin_ult1_10month_ago:** ind_viv_fin_ult1 ownership 10 months ago.
**ind_nomina_ult1_10month_ago:** ind_nomina_ult1 ownership 10 months ago.
**ind_nom_pens_ult1_10month_ago:** ind_nom_pens_ult1 ownership 10 months ago.
**ind_recibo_ult1_10month_ago:** ind_recibo_ult1 ownership 10 months ago.

**ind_ahor_fin_ult1_11month_ago:** ind_ahor_fin_ult1 ownership 11 months ago.
**ind_aval_fin_ult1_11month_ago:** ind_aval_fin_ult1 ownership 11 months ago.

**ind_cco_fin_ult1_11month_ago:** ind_cco_fin_ult1 ownership 11 months ago.
**ind_cder_fin_ult1_11month_ago:** ind_cder_fin_ult1 ownership 11 months ago.
**ind_cno_fin_ult1_11month_ago:** ind_cno_fin_ult1 ownership 11 months ago.
**ind_ctju_fin_ult1_11month_ago:** ind_ctju_fin_ult1 ownership 11 months ago.
**ind_ctma_fin_ult1_11month_ago:** ind_ctma_fin_ult1 ownership 11 months ago.
**ind_ctop_fin_ult1_11month_ago:** ind_ctop_fin_ult1 ownership 11 months ago.
**ind_ctpp_fin_ult1_11month_ago:** ind_ctpp_fin_ult1 ownership 11 months ago.
**ind_deco_fin_ult1_11month_ago:** ind_deco_fin_ult1 ownership 11 months ago.
**ind_deme_fin_ult1_11month_ago:** ind_deme_fin_ult1 ownership 11 months ago.
**ind_dela_fin_ult1_11month_ago:** ind_dela_fin_ult1 ownership 11 months ago.
**ind_ecue_fin_ult1_11month_ago:** ind_ecue_fin_ult1 ownership 11 months ago.
**ind_fond_fin_ult1_11month_ago:** ind_fond_fin_ult1 ownership 11 months ago.
**ind_hip_fin_ult1_11month_ago:** ind_hip_fin_ult1 ownership 11 months ago.
**ind_plan_fin_ult1_11month_ago:** ind_plan_fin_ult1 ownership 11 months ago.
**ind_pres_fin_ult1_11month_ago:** ind_pres_fin_ult1 ownership 11 months ago.
**ind_reca_fin_ult1_11month_ago:** ind_reca_fin_ult1 ownership 11 months ago.
**ind_tjcr_fin_ult1_11month_ago:** ind_tjcr_fin_ult1 ownership 11 months ago.
**ind_valo_fin_ult1_11month_ago:** ind_valo_fin_ult1 ownership 11 months ago.
**ind_viv_fin_ult1_11month_ago:** ind_viv_fin_ult1 ownership 11 months ago.
**ind_nomina_ult1_11month_ago:** ind_nomina_ult1 ownership 11 months ago.
**ind_nom_pens_ult1_11month_ago:** ind_nom_pens_ult1 ownership 11 months ago.
**ind_recibo_ult1_11month_ago:** ind_recibo_ult1 ownership 11 months ago.

Months since last owned: For each product we derived a feature which represented the number of months since that product has been owned. If the product is not owned, then it is given a default value of 999. Following are the features derived using this approach.

**ind_ahor_fin_ult1_last_owned:** Number of months since ind_ahor_fin_ult1 is owned.
**ind_aval_fin_ult1_last_owned:** Number of months since ind_aval_fin_ult1 is owned.
**ind_cco_fin_ult1_last_owned:** Number of months since ind_cco_fin_ult1 is owned.
**ind_cder_fin_ult1_last_owned:** Number of months since ind_cder_fin_ult1 is owned.
**ind_cno_fin_ult1_last_owned:** Number of months since ind_cno_fin_ult1 is owned.
**ind_ctju_fin_ult1_last_owned:** Number of months since ind_ctju_fin_ult1 is owned.
**ind_ctma_fin_ult1_last_owned:** Number of months since ind_ctma_fin_ult1 is owned.
**ind_ctop_fin_ult1_last_owned:** Number of months since ind_ctop_fin_ult1 is owned.
**ind_ctpp_fin_ult1_last_owned:** Number of months since ind_ctpp_fin_ult1 is owned.
**ind_deco_fin_ult1_last_owned:** Number of months since ind_deco_fin_ult1 is owned.
**ind_deme_fin_ult1_last_owned:** Number of months since ind_deme_fin_ult1 is owned.
**ind_dela_fin_ult1_last_owned:** Number of months since ind_dela_fin_ult1 is owned.
**ind_ecue_fin_ult1_last_owned:** Number of months since ind_ecue_fin_ult1 is owned.
**ind_fond_fin_ult1_last_owned:** Number of months since ind_fond_fin_ult1 is owned.
**ind_hip_fin_ult1_last_owned:** Number of months since ind_hip_fin_ult1 is owned.

**ind_plan_fin_ult1_last_owned:** Number of months since ind_plan_fin_ult1 is owned.
**ind_pres_fin_ult1_last_owned:** Number of months since ind_pres_fin_ult1 is owned.
**ind_reca_fin_ult1_last_owned:** Number of months since ind_reca_fin_ult1 is owned.
**ind_tjcr_fin_ult1_last_owned:** Number of months since ind_tjcr_fin_ult1 is owned.
**ind_valo_fin_ult1_last_owned:** Number of months since ind_valo_fin_ult1 is owned.
**ind_viv_fin_ult1_last_owned:** Number of months since ind_viv_fin_ult1 is owned.
**ind_nomina_ult1_last_owned:** Number of months since ind_nomina_ult1 is owned.
**ind_nom_pens_ult1_last_owned:** Number of months since ind_nom_pens_ult1 is owned. **ind_recibo_ult1_last_owned:** Number of months since ind_recibo_ult1 is owned. **total_products:** Total number of products owned previous month.

Windows of product ownership: For each window size look back at previous months and see if the product was ever owned. I did this by adding the value of the ownership variable X months ago for X = 1: window_size. Then converting to a binary indicator if the value is positive. I am using a using size of 2 to 6. Following are the features derived using this method:

**ind_ahor_fin_ult1_owned_within_2months:** Is ind_ahor_fin_ult1 owned within last two months
**ind_ahor_fin_ult1_owned_within_3months:** Is ind_ahor_fin_ult1 owned within last three months
**ind_ahor_fin_ult1_owned_within_4months:** Is ind_ahor_fin_ult1 owned within last four months
**ind_ahor_fin_ult1_owned_within_5months:** Is ind_ahor_fin_ult1 owned within last five months
**ind_ahor_fin_ult1_owned_within_6months:** Is ind_ahor_fin_ult1 owned within last six months
**ind_aval_fin_ult1_owned_within_2months:** Is ind_aval_fin_ult1 owned within last two months.
**ind_aval_fin_ult1_owned_within_3months:** Is ind_aval_fin_ult1 owned within last three months.
**ind_aval_fin_ult1_owned_within_4months:** Is ind_aval_fin_ult1 owned within last four months.
**ind_aval_fin_ult1_owned_within_5months:** Is ind_aval_fin_ult1 owned within last five months.
**ind_aval_fin_ult1_owned_within_6months:** Is ind_aval_fin_ult1 owned within last six months.
**ind_cco_fin_ult1_owned_within_2months:** Is ind_cco_fin_ult1 owned within last two months.
**ind_cco_fin_ult1_owned_within_3months:** Is ind_cco_fin_ult1 owned within last three months.
**ind_cco_fin_ult1_owned_within_4months:** Is ind_cco_fin_ult1 owned within last four months.
**ind_cco_fin_ult1_owned_within_5months:** Is ind_cco_fin_ult1 owned within last five months.
**ind_cco_fin_ult1_owned_within_6months:** Is ind_cco_fin_ult1 owned within last six months.
**ind_cder_fin_ult1_owned_within_2months:** Is ind_cder_fin_ult1 owned within last two

months.
**ind_cder_fin_ult1_owned_within_3months:** Is ind_cder_fin_ult1 owned within last three months.
**ind_cder_fin_ult1_owned_within_4months:** Is ind_cder_fin_ult1 owned within last four months.
**ind_cder_fin_ult1_owned_within_5months:** Is ind_cder_fin_ult1 owned within last five months.
**ind_cder_fin_ult1_owned_within_6months:** Is ind_cder_fin_ult1 owned within last six months.
**ind_cno_fin_ult1_owned_within_2months:** Is ind_cno_fin_ult1 owned within last two months.
**ind_cno_fin_ult1_owned_within_3months:** Is ind_cno_fin_ult1 owned within last three months.
**ind_cno_fin_ult1_owned_within_4months:** Is ind_cno_fin_ult1 owned within last four months.
**ind_cno_fin_ult1_owned_within_5months:** Is ind_cno_fin_ult1 owned within last five months.
**ind_cno_fin_ult1_owned_within_6months:** Is ind_cno_fin_ult1 owned within last six months.
**ind_ctju_fin_ult1_owned_within_2months:** Is ind_ctju_fin_ult1 owned within last two months.
**ind_ctju_fin_ult1_owned_within_3months:** Is ind_ctju_fin_ult1 owned within last three months.
**ind_ctju_fin_ult1_owned_within_4months:** Is ind_ctju_fin_ult1 owned within last four months.
**ind_ctju_fin_ult1_owned_within_5months:** Is ind_ctju_fin_ult1 owned within last five months.
**ind_ctju_fin_ult1_owned_within_6months:** Is ind_ctju_fin_ult1 owned within last six months.
**ind_ctma_fin_ult1_owned_within_2months:** Is ind_ctma_fin_ult1 owned within last two months.
**ind_ctma_fin_ult1_owned_within_3months:** Is ind_ctma_fin_ult1 owned within last three months.
**ind_ctma_fin_ult1_owned_within_4months:** Is ind_ctma_fin_ult1 owned within last four months.
**ind_ctma_fin_ult1_owned_within_5months:** Is ind_ctma_fin_ult1 owned within last five months.
**ind_ctma_fin_ult1_owned_within_6months:** Is ind_ctma_fin_ult1 owned within last six months.
**ind_ctop_fin_ult1_owned_within_2months:** Is ind_ctop_fin_ult1 owned within last two months.
**ind_ctop_fin_ult1_owned_within_3months:** Is ind_ctop_fin_ult1 owned within last three months.
**ind_ctop_fin_ult1_owned_within_4months:** Is ind_ctop_fin_ult1 owned within last four months.
**ind_ctop_fin_ult1_owned_within_5months:** Is ind_ctop_fin_ult1 owned within last five months.
**ind_ctop_fin_ult1_owned_within_6months:** Is ind_ctop_fin_ult1 owned within last six months.
**ind_ctpp_fin_ult1_owned_within_2months:** Is ind_ctpp_fin_ult1 owned within last two months.
**ind_ctpp_fin_ult1_owned_within_3months:** Is ind_ctpp_fin_ult1 owned within last three

months.
**ind_ctpp_fin_ult1_owned_within_4months:** Is ind_ctpp_fin_ult1 owned within last four months.
**ind_ctpp_fin_ult1_owned_within_5months:** Is ind_ctpp_fin_ult1 owned within last five months.
**ind_ctpp_fin_ult1_owned_within_6months:** Is ind_ctpp_fin_ult1 owned within last six months.
**ind_deco_fin_ult1_owned_within_2months:** Is ind_deco_fin_ult1 owned within last two months.
**ind_deco_fin_ult1_owned_within_3months:** Is ind_deco_fin_ult1 owned within last three months.
**ind_deco_fin_ult1_owned_within_4months:** Is ind_deco_fin_ult1 owned within last four months.
**ind_deco_fin_ult1_owned_within_5months:** Is ind_deco_fin_ult1 owned within last five months.
**ind_deco_fin_ult1_owned_within_6months:** Is ind_deco_fin_ult1 owned within last six months.
**ind_deme_fin_ult1_owned_within_2months:** Is ind_deme_fin_ult1 owned within last two months.
**ind_deme_fin_ult1_owned_within_3months:** Is ind_deme_fin_ult1 owned within last three months.
**ind_deme_fin_ult1_owned_within_4months:** Is ind_deme_fin_ult1 owned within last four months.
**ind_deme_fin_ult1_owned_within_5months:** Is ind_deme_fin_ult1 owned within last five months.
**ind_deme_fin_ult1_owned_within_6months:** Is ind_deme_fin_ult1 owned within last six months.
**ind_dela_fin_ult1_owned_within_2months**: Is ind_dela_fin_ult1 owned within last two months.
**ind_dela_fin_ult1_owned_within_3months**: Is ind_dela_fin_ult1 owned within last three months.
**ind_dela_fin_ult1_owned_within_4months**: Is ind_dela_fin_ult1 owned within last four months.
**ind_dela_fin_ult1_owned_within_5months**: Is ind_dela_fin_ult1 owned within last five months.
**ind_dela_fin_ult1_owned_within_6months**: Is ind_dela_fin_ult1 owned within last six months.
**ind_ecue_fin_ult1_owned_within_2months:** Is ind_ecue_fin_ult1 owned within last two months.
**ind_ecue_fin_ult1_owned_within_3months:** Is ind_ecue_fin_ult1 owned within last three months.
**ind_ecue_fin_ult1_owned_within_4months:** Is ind_ecue_fin_ult1 owned within last four months.
**ind_ecue_fin_ult1_owned_within_5months:** Is ind_ecue_fin_ult1 owned within last five

months.
**ind_ecue_fin_ult1_owned_within_6months:** Is ind_ecue_fin_ult1 owned within last six months.
**ind_fond_fin_ult1_owned_within_2months:** Is ind_fond_fin_ult1 owned within last two months.
**ind_fond_fin_ult1_owned_within_3months:** Is ind_fond_fin_ult1 owned within last three months.
**ind_fond_fin_ult1_owned_within_4months:** Is ind_fond_fin_ult1 owned within last four months.
**ind_fond_fin_ult1_owned_within_5months:** Is ind_fond_fin_ult1 owned within last five months.
**ind_fond_fin_ult1_owned_within_6months:** Is ind_fond_fin_ult1 owned within last six months.
**ind_hip_fin_ult1_owned_within_2months:** Is ind_hip_fin_ult1 owned within last two months.
**ind_hip_fin_ult1_owned_within_3months:** Is ind_hip_fin_ult1 owned within last three months.
**ind_hip_fin_ult1_owned_within_4months:** Is ind_hip_fin_ult1 owned within last four months.
**ind_hip_fin_ult1_owned_within_5months:** Is ind_hip_fin_ult1 owned within last five months.
**ind_hip_fin_ult1_owned_within_6months:** Is ind_hip_fin_ult1 owned within last six months.
**ind_plan_fin_ult1_owned_within_2months:** Is ind_plan_fin_ult1 owned within last two months.
**ind_plan_fin_ult1_owned_within_3months:** Is ind_plan_fin_ult1 owned within last three months.
**ind_plan_fin_ult1_owned_within_4months:** Is ind_plan_fin_ult1 owned within last four months.
**ind_plan_fin_ult1_owned_within_5months:** Is ind_plan_fin_ult1 owned within last five months.
**ind_plan_fin_ult1_owned_within_6months:** Is ind_plan_fin_ult1 owned within last six months.
**ind_pres_fin_ult1_owned_within_2months:** Is ind_pres_fin_ult1 owned within last two months.
**ind_pres_fin_ult1_owned_within_3months:** Is ind_pres_fin_ult1 owned within last three months.
**ind_pres_fin_ult1_owned_within_4months:** Is ind_pres_fin_ult1 owned within last four months.
**ind_pres_fin_ult1_owned_within_5months:** Is ind_pres_fin_ult1 owned within last five months.
**ind_pres_fin_ult1_owned_within_6months:** Is ind_pres_fin_ult1 owned within last six months.
**ind_reca_fin_ult1_owned_within_2months:** Is ind_reca_fin_ult1 owned within last two months.
**ind_reca_fin_ult1_owned_within_3months:** Is ind_reca_fin_ult1 owned within last three months.
**ind_reca_fin_ult1_owned_within_4months:** Is ind_reca_fin_ult1 owned within last four months.
**ind_reca_fin_ult1_owned_within_5months:** Is ind_reca_fin_ult1 owned within last five

months.
**ind_reca_fin_ult1_owned_within_6months:** Is ind_reca_fin_ult1 owned within last six months.
**ind_tjcr_fin_ult1_owned_within_2months**: is ind_tjcr_fin_ult1 owned within last two months.
**ind_tjcr_fin_ult1_owned_within_3months**: is ind_tjcr_fin_ult1 owned within last three months.
**ind_tjcr_fin_ult1_owned_within_4months**: is ind_tjcr_fin_ult1 owned within last four months.
**ind_tjcr_fin_ult1_owned_within_5months**: is ind_tjcr_fin_ult1 owned within last five months. months.
**ind_tjcr_fin_ult1_owned_within_6months**: is ind_tjcr_fin_ult1 owned within last six months.
**ind_valo_fin_ult1_owned_within_2months:** is ind_valo_fin_ult1 owned within last two months.
**ind_valo_fin_ult1_owned_within_3months:** is ind_valo_fin_ult1 owned within last three months.
**ind_valo_fin_ult1_owned_within_4months:** is ind_valo_fin_ult1 owned within last four months.
**ind_valo_fin_ult1_owned_within_5months:** is ind_valo_fin_ult1 owned within last five months.
**ind_valo_fin_ult1_owned_within_6months:** is ind_valo_fin_ult1 owned within last six months.
**ind_viv_fin_ult1_owned_within_2months**: is ind_viv_fin_ult1 owned within last two months.
**ind_viv_fin_ult1_owned_within_3months**: is ind_viv_fin_ult1 owned within last three months.
**ind_viv_fin_ult1_owned_within_4months:** is ind_viv_fin_ult1 owned within last four months.
**ind_viv_fin_ult1_owned_within_5months:** is ind_viv_fin_ult1 owned within last five months.
**ind_viv_fin_ult1_owned_within_6months:** is ind_viv_fin_ult1 owned within last six months.
**ind_nomina_ult1_owned_within_2months:** is ind_nomina_ult1 owned within last two months.
**ind_nomina_ult1_owned_within_3months:** is ind_nomina_ult1 owned within last three months.
**ind_nomina_ult1_owned_within_4months:** is ind_nomina_ult1 owned within last four months.
**ind_nomina_ult1_owned_within_5months:** is ind_nomina_ult1 owned within last five months.
**ind_nomina_ult1_owned_within_6months:** is ind_nomina_ult1 owned within last six months.
**ind_nom_pens_ult1_owned_within_2months:** is ind_nom_pens_ult1 owned within last two months.
**ind_nom_pens_ult1_owned_within_3months:** is ind_nom_pens_ult1 owned within last three months.
**ind_nom_pens_ult1_owned_within_4months:** is ind_nom_pens_ult1 owned within last four months.
**ind_nom_pens_ult1_owned_within_5months:** is ind_nom_pens_ult1 owned within last five months.
**ind_nom_pens_ult1_owned_within_6months:** is ind_nom_pens_ult1 owned within last six months.
**ind_recibo_ult1_owned_within_2months:** is ind_recibo_ult1 owned within last two months.
**ind_recibo_ult1_owned_within_3months:** is ind_recibo_ult1 owned within last three months.
**ind_recibo_ult1_owned_within_4months:** is ind_recibo_ult1 owned within last four months.

**ind_recibo_ult1_owned_within_5months:** is ind_recibo_ult1 owned within last five months.
**ind_recibo_ult1_owned_within_6months:** is ind_recibo_ult1 owned within last six months.

**3. Data Analysis**

After the completion of data preprocessing, we will perform preliminary investigations on data. We will be working on a dataset for generating various plots to extract and explore insights hidden in our data.

1. **The first day of the week the customer signed a contract with the bank, number of customers that became first holder by day of week.**

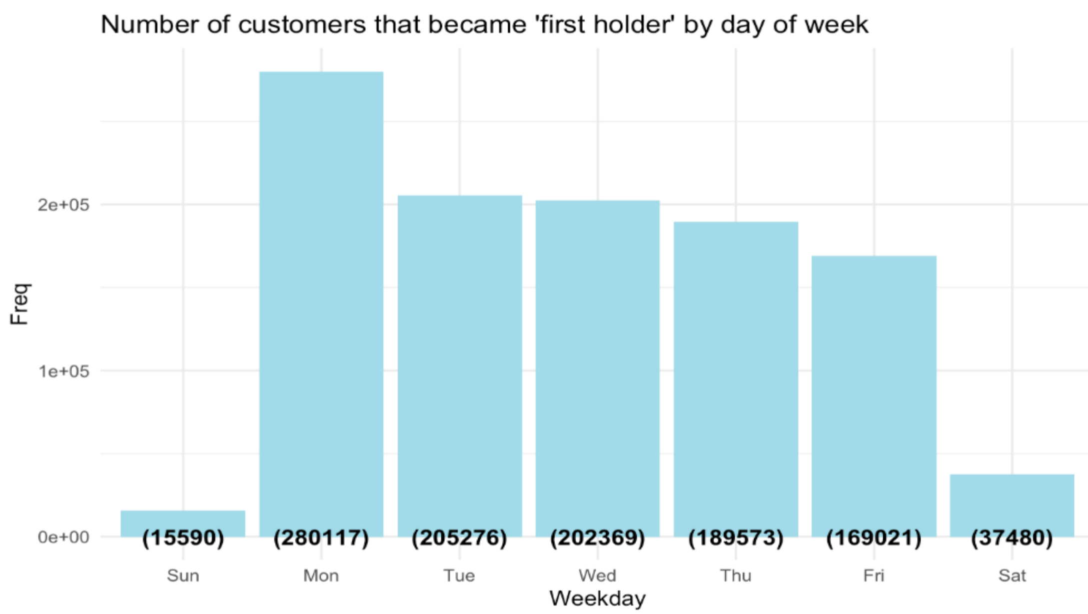



**Figure 3a**

Figure 3a shows the frequency of customers that became ‘first holder’ by day of the week. Moreover, it shows the first day of the week the customer signed a contract with the bank. From the above observation out of all the days, “Monday” has the highest frequency of the customer who became ‘first holder’. It could also be inferred as the day goes by and as we move closer to the weekend the frequency of the customer that became first holder decreases, Sunday having the lowest frequency.

2. **Year & Month (year and month the customer first signed a contract with the bank)**

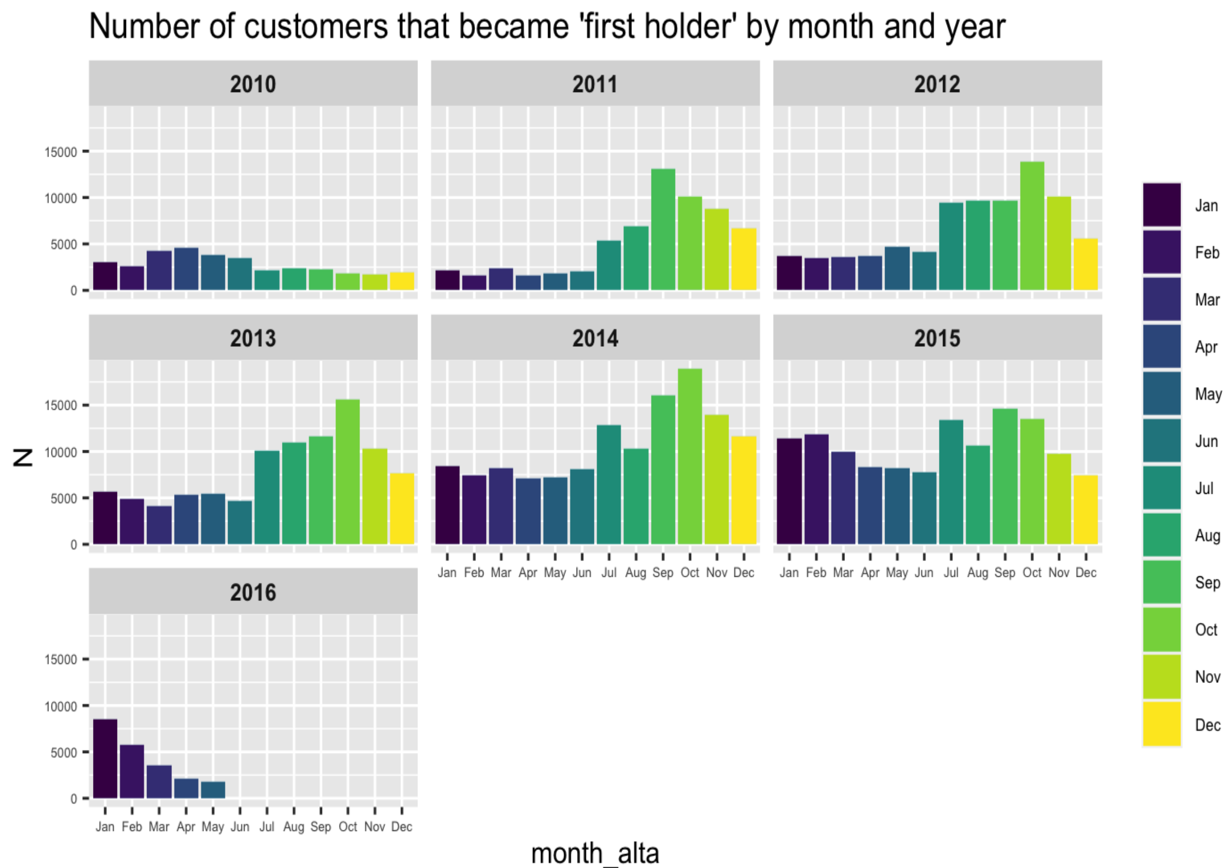



**Figure 3b**

This graph shows the frequency of customers who became 'first holders' by month and year. From the above observation we can see the highest frequency of customers who became 'fist holder' in the year "2014" for October month. However, it is lowest for "2016" in the month of May. Moreover, for September and October for the following year: "2011", "2012", "2013", "2014" and "2015" it is pretty much consistent and above "10000". For the year "2010", there is no major difference as all months have an almost consistent frequency of customers not more than "5000". Also, "2016" can be considered an exception as after May no customers became 'first holder' for that particular year.

### 3. Service Changes by Month

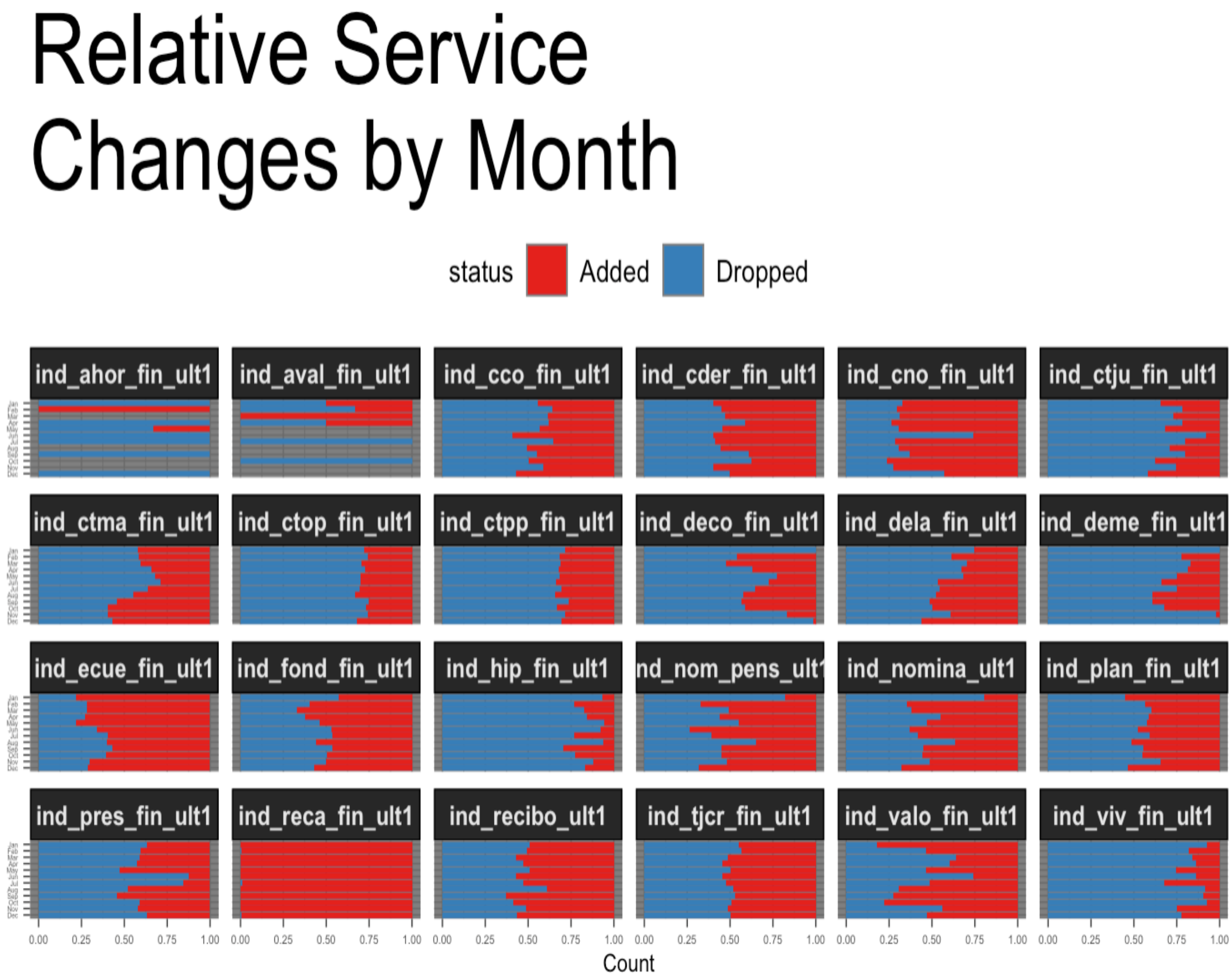



**Figure 3c**

Figure 3c shows how the ratio of financial products (services) changes by month. Here, “ind_ahor_fin_ult1” has the highest change of service in the month of ‘January’ and this product has the least number of added services among all products. The other service change ratio is seen in the month of May up to 65% and there is no add or drop-in service for the following months: March, August, October, and November for the same financial product. Also, for the “ind_reca_fin_ult1”, there is consistent addition of services for all months. However, the higher drop-in services can be seen for the “ind_deme_fin_ult1” and “ind_deco_fin_ult1” in the months of ‘January’ and ‘December’. Similarly for “ind_hip_fin_ult1” and “ ind_viv_fin_ult1”  there is a maximum drop in service in the month of 'January' up to 95% . Moreover, there are financial services like "ind_tjcr_fin_ult1" and "ind_recibo_ult1" which have almost equal amounts of change in status.

**4. Average Service Changes by Month**

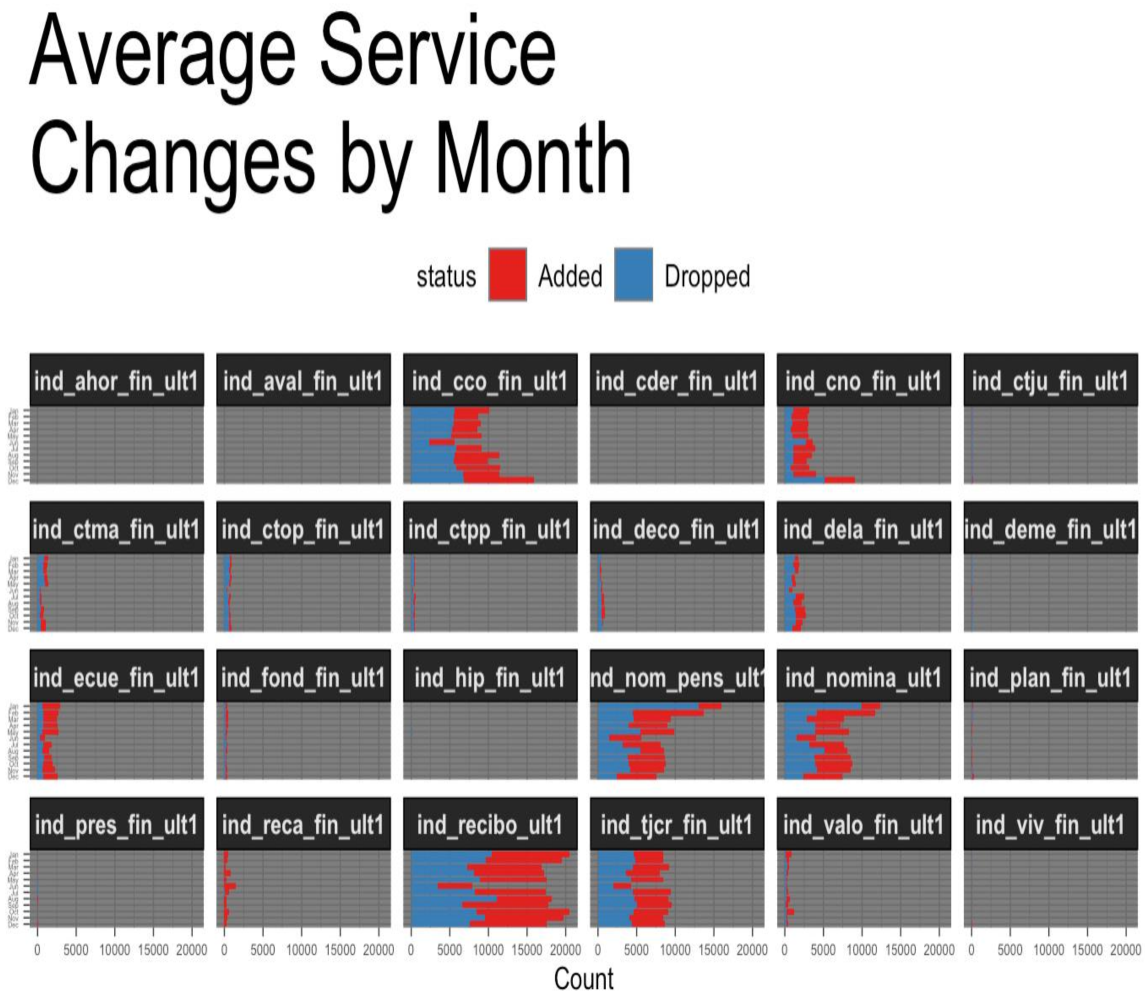



**Figure 3d**

Here, for the "ind_ahor_fin_ult1" and "ind_aval_fin_ult1" there is no change in average service. However, maximum change in service (add and drop) is for "ind_recibo_ult1" and highest in addition is in the month of September and maximum dropping of service is in month of 'July'.

## 5. Service Changes by Gender


**Figure 3e**

We will do Gender-based analysis to improve the efficiency and impact of financial services, to make them more inclusive and responsive to the needs of all concerned customers and to reduce inequalities between the Genders. Here, for figure 3(e) we can see that the graph is representing change in services associated with gender. For the "ind_cco_fin_ult1" the add and drop-in service is almost nearby for gender "V" and for "H" it is less than V but close in change. The least number of changes are almost consistent for "ind_hip_fin_ult1", "ind_fond_fin_ult1", "ind_ctpp_fin_ult1"," ind_deco_fin_ult1"," ind_plan_fin_ult1", "ind_reca_fin_ult1", "ind_valo_fin_ult1". Whereas the highest change in service ratio is for "ind_recibo_ult1", "V" has above 150000 and "H" have till 50000.

6. **Seniority Vs Age**
By doing Seniority Vs Age analysis of Customers their status, rank, or precedence can be known.

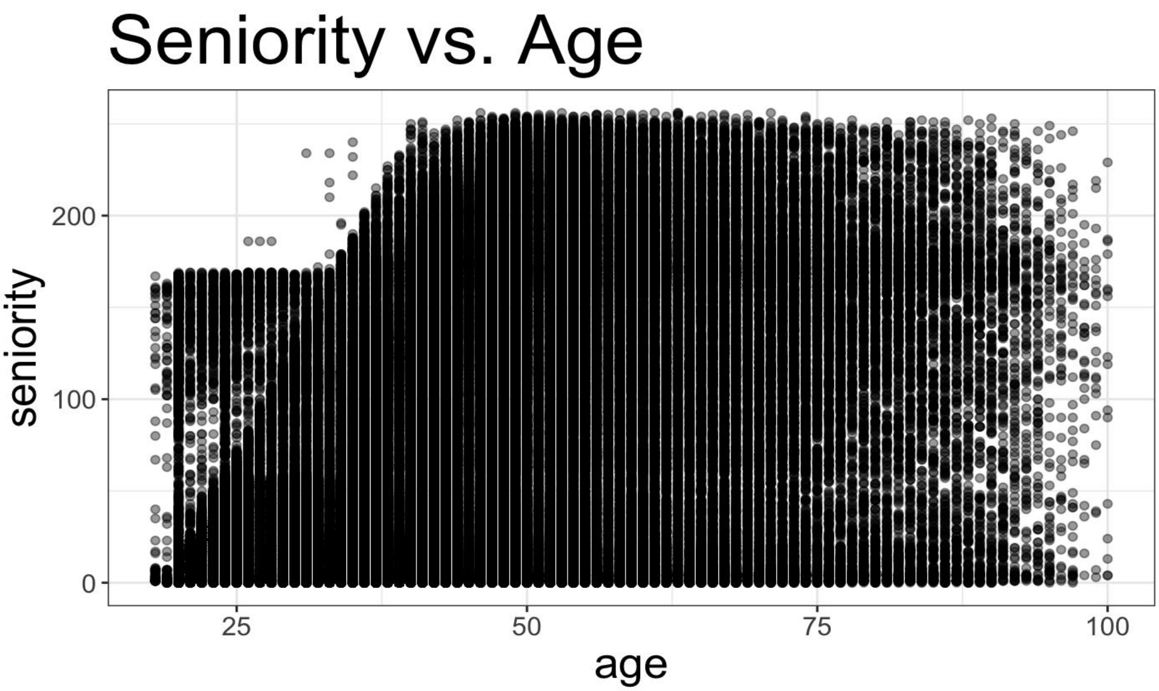



**Figure 3f**

The graph represents the customers seniority and customers' age. As shown in the figure 3(f), the seniority remains consistent till age of 37 and keeps on rising till 43 and remains consistent up to a certain age and then slightly decreases till 100.

## 7. Service change by City

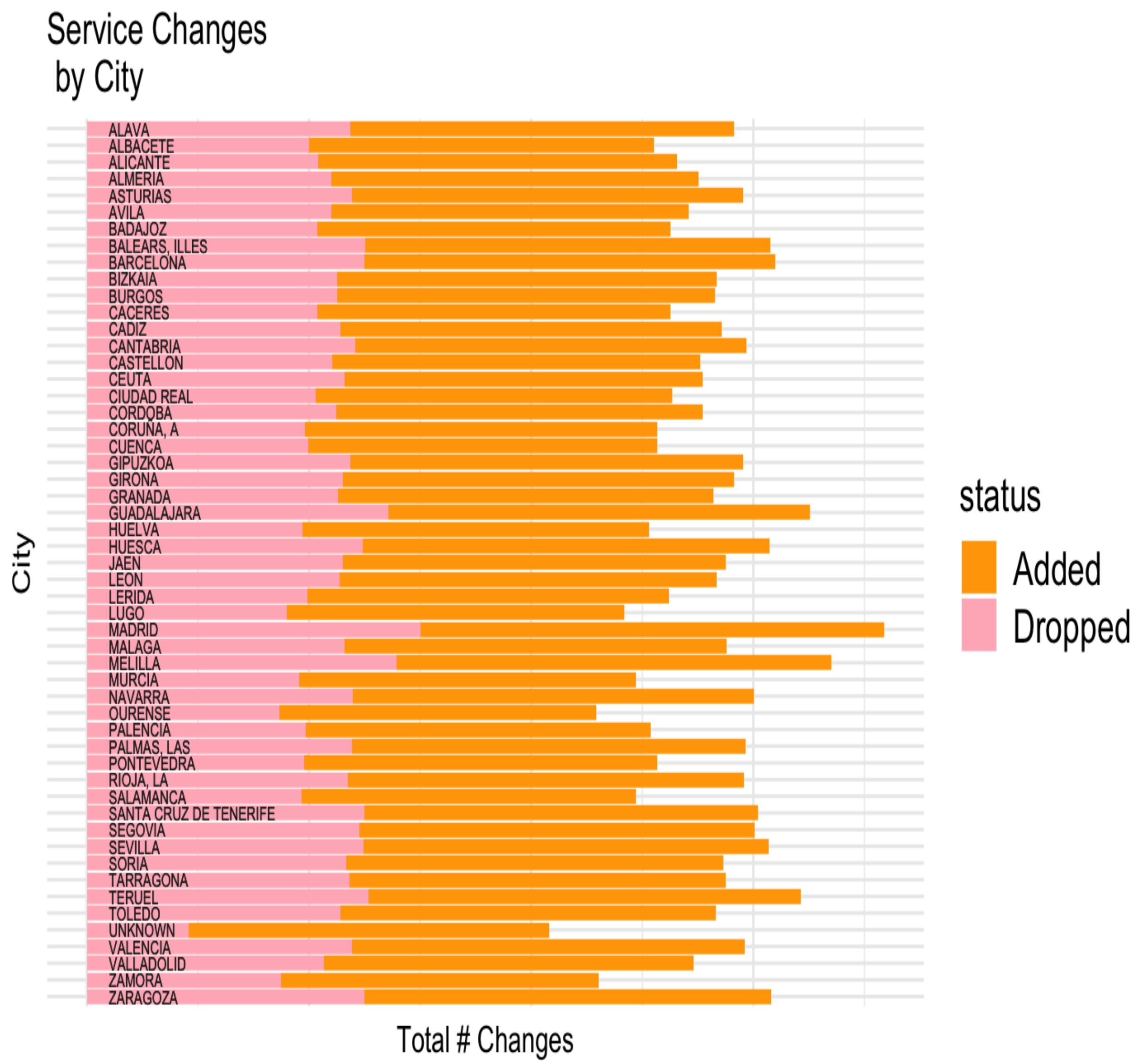



**Figure 3g**

We will determine the total number of changes in Service. According to the above graph the highest number of services added to the city “Madrid”. Almost all the cities have greater number of service additions compared to drop-in services.

**8. Income Vs Age**

Using this information, we can extract demographic information about the customers.

Over here, the income is gradually decreasing as the age is decreasing and the addition and drop of income ratio is almost consistent.

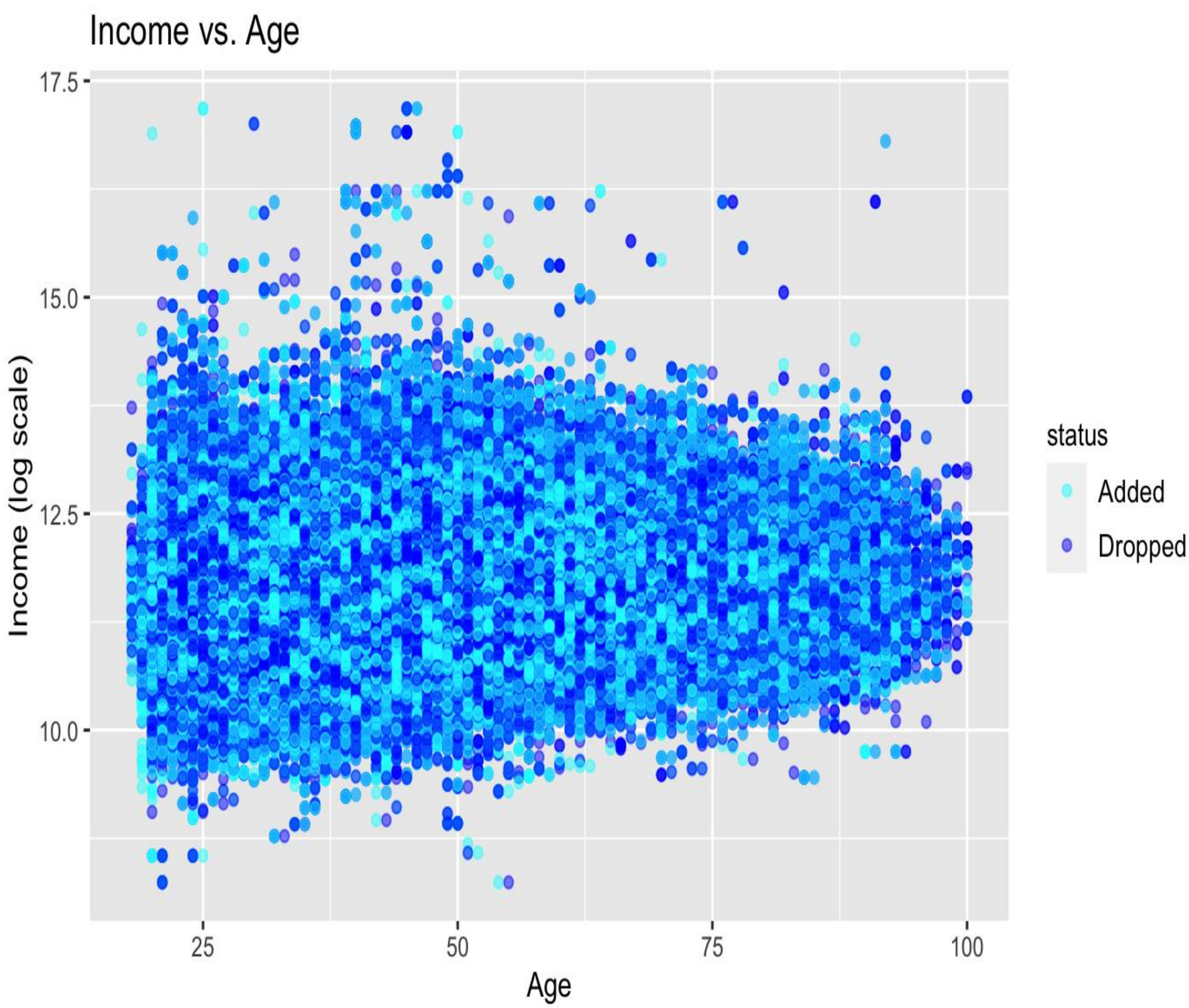



**Figure 3h**

**9.Service changes by seniority**

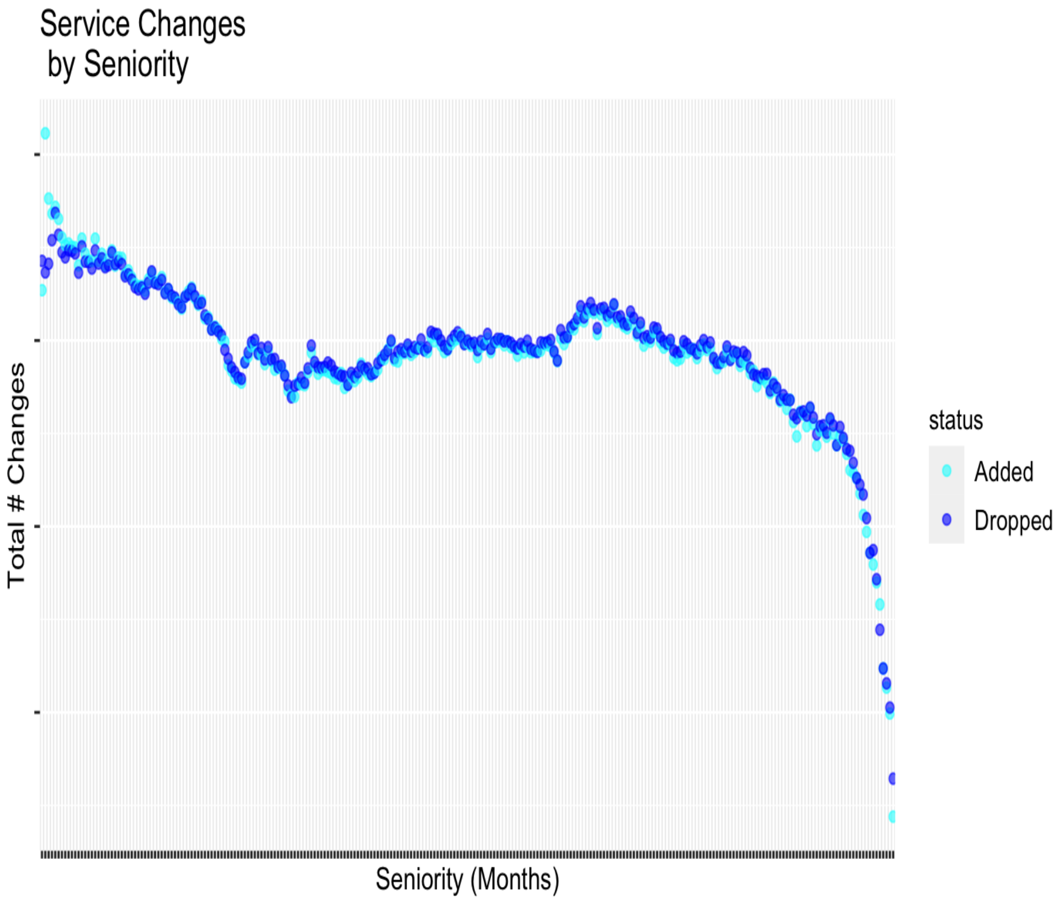



**Figure 3i**

From the figure3(i) it can be inferred that the Total number of changes in service is decreasing with the seniority in respect to months and remains consistent for a couple of months, then again decreases. Moreover, drop-in service is more compared to addition.

**10. Normalized service changes by new status**

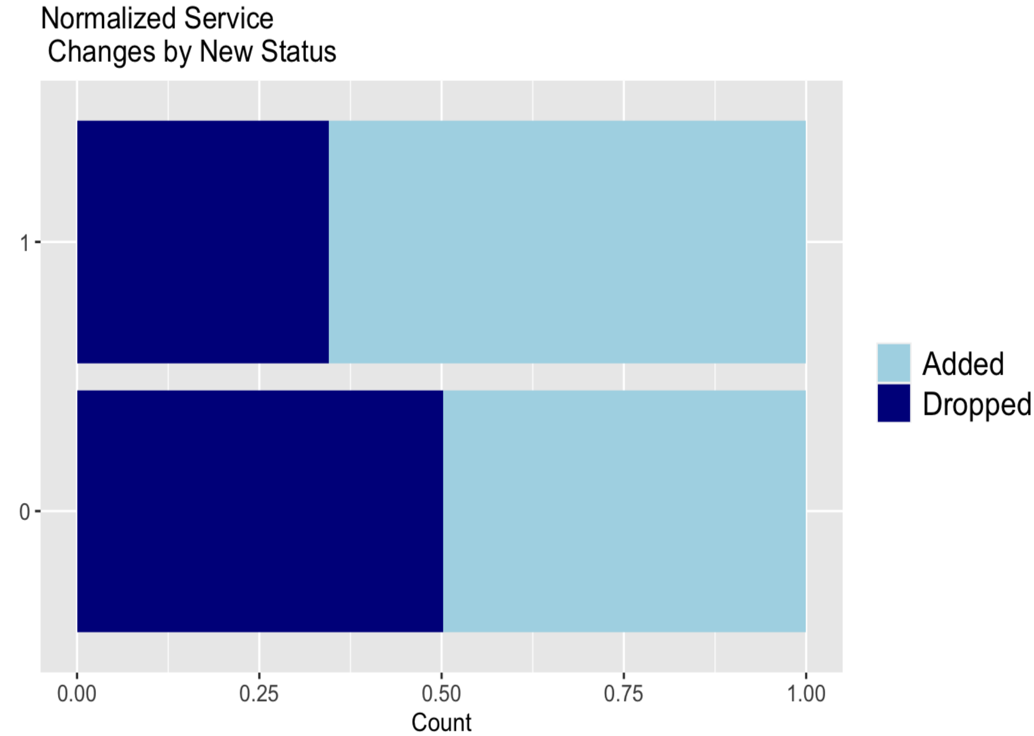



**Figure 3j**

From the above plot it can be seen that the customers which are new have index 1 if the customer registered in the last 6 months else 0. For the new customers the count of services added are more up to 65% while the number of services dropped are less up to 33%. However, for the customers who are not new and registered before 6 months have a 50% - 50% ratio of addition and drop-in Service.

**11. Age distributed according to groups.**

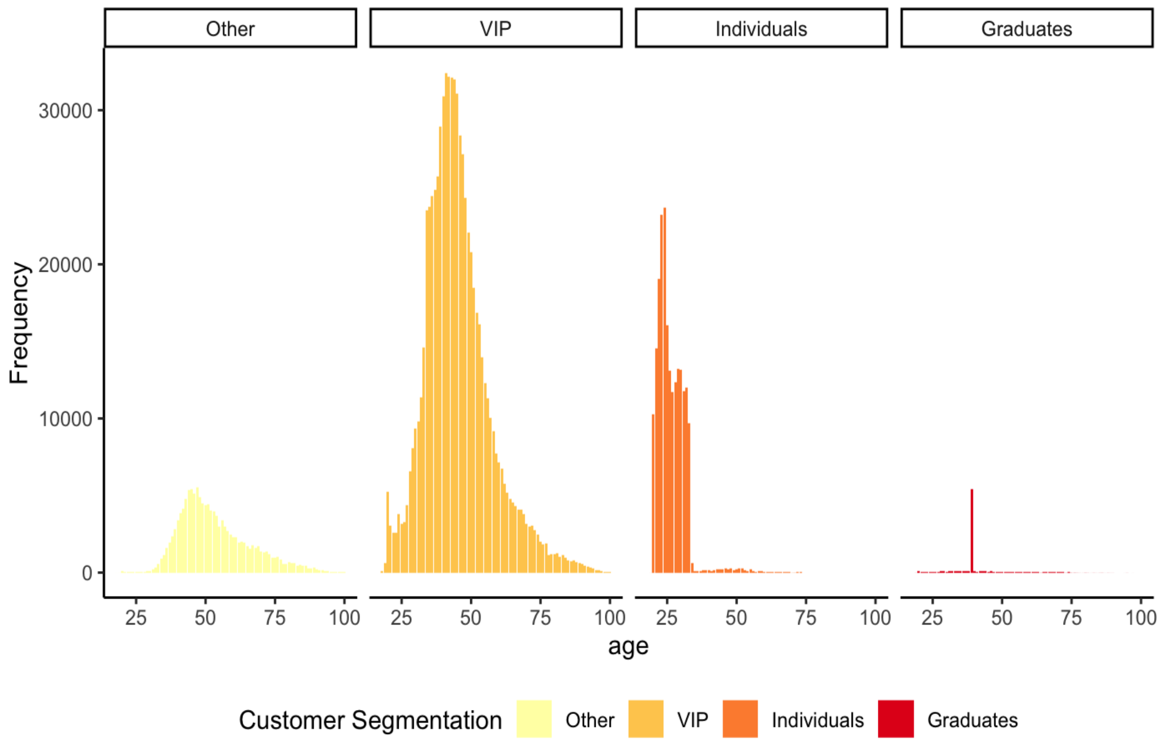



**Figure 3k**

Histogram reveals that the younger group belongs to college graduates and the older group belongs to an individual group of customers. It can be interpreted that people who are near the age of 50 have higher frequency and belong to VIP groups. Similarly, those who are at 25 and have the higher frequency are considered to be individuals. Moreover, the highest frequency is seen in VIP groups whereas lowest number of frequencies is seen in college graduates.

### 4. Model Training, Validation, and Conclusion

We have built an ensemble of two models – xgboost: multiclass and xgboost: binary. We have performed same preprocessing for both the models which as follows:

**Step1:** We used 5th and 11th month for training and 17th month for validation. After that we used 6th and 12th month for training and 18th month for test.

**Step2:** Features used for modelling:

**Numeric Features used for modelling:**
age, renta, antiguedad, 24 derived features that end with count( example, ind_ahor_fin_ult1_purchase_count), total_products, num_transactions, num_purchases_1_months_ago, num_purchases_2_months_ago, num_purchases_3_months_ago, num_purchases_4_months_ago.

**Categorical Features to use for modelling:**
sexo, ind_nuevo, ind_empleado, segmento, nomprov, indext, indresi, indrel, tiprel_1mes, 264 derived features that end with month_ago( example, ind_ahor_fin_ult1_1month_ago), 120 derived features that contain owned_within in its name (example, ind_ahor_fin_ult1_owned_within_2months).

**Creation of Dummy Variables:** We have created dummy variables for each of the above categorical variables.

**Model1: xgboost: Multiclass**
**Parameters used:**

We ran a grid search on the following values of depth and learning rate.
depth: 3, 5, 7, 9, 11, 15
learning rate: 0.01,0.025, 0.05,0.1,0.25,0.5
The best combination turned out to be: depth=7, learning rate=0.05
Other parameters are:

objective: "multi:softprob"
nthread = 4
nround = 175
num_class=22

**Validation score MAP@7 = 0.889565471049386**

**Output Files:** xgboost_preds_val_future_multiclass.csv
xgboost_preds_test_multiclass.csv

**Model2: xgboost: single-class**
**Parameters used:**

We ran a grid search on the following values of depth and learning rate.
depth: 3, 5, 7, 9, 11, 15
learning rate: 0.01,0.025, 0.05,0.1,0.25,0.5

The best combination turned out to be: depth=7, learning rate=0.05
Other parameters are:
objective: " binary:logistic"
nthread = 4
nround = 80

**"Validation score MAP@7 = 0.89275133654992"**

**Output Files:** xgboost_preds_val_future_singleclass.csv
xgboost_preds_test_singleclass.csv

**Combining Predictions from both the models:**

We give a weight of .1 to multi-class model whereas 1 to single-class model.

**Output Files:**
combined_preds_val.csv
combined_preds_test.csv

Let's analyze a few records from our validation set(combined_preds_val.csv) to verify if model predictions are like actual products:

ncodpers: 15889
Actual Products: ind_cco_fin_ult1, ind_ctpp_fin_ult1, ind_valo_fin_ult1
Predicted Products: ind_tjcr_fin_ult1, ind_valo_fin_ult1, ind_cco_fin_ult1, ind_ctpp_fin_ult1

Prediction made by the ensemble contain one extra product ind_tjcr_fin_ult1. All the remaining predictions are correct.

ncodpers: 15929
Actual Products: ind_cco_fin_ult1, ind_ctpp_fin_ult1, ind_ecue_fin_ult1, ind_tjcr_fin_ult1, ind_valo_fin_ult1

Predicted Products: ind_cco_fin_ult1, ind_ctpp_fin_ult1, ind_ecue_fin_ult1, ind_tjcr_fin_ult1, ind_valo_fin_ult1, ind_recibo_ult1

Prediction made by the ensemble contain one extra product ind_recibo_ult1. All the remaining predictions are correct.

Final recommendations file: recommendations_xgboost.csv

**5. Conclusion and Future Work:**

Ensemble is correctly predicting **True Positives**. However, it is also producing some **False Positives**. Since all the True Positives are captured correctly, the marketing team might give some lubricative offers to its customers to delight them and reduce their churn.

For future scope of the project, we would like to leverage collaborative filtering to build the recommendation system.

**6. Data Sources:**

All the data files for this competition is kept at below location:

https://www.kaggle.com/competitions/santander-product-recommendation/data

**7. Source Code and Files:**

https://drive.google.com/drive/folders/1ul0pwbqiSu58jxblKHu_FuqN43kneXEh?usp=share_link

Libraries: data.table, plyr, tidyr, lubridate, ggplot2, fasttime, xgboost 1.1.1.1, caret, pROC

**8. Bibliography**

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016.

Schafer, J. Ben, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative filtering recommender systems." In *The adaptive web*, pp. 291-324. Springer, Berlin, Heidelberg, 2007.

Leung, K. Ming. "Naive bayesian classifier." *Polytechnic University Department of Computer Science/Finance and Risk Engineering* 2007 (2007): 123-156.