Grupo Bimbo Demand Prediction

Presented by

Group 3
Napattarapon Pranmontri 5930274821
Patawee Pingclasai 5930304021
Date: 7 January 2020
ENG4 Building Room 409

Grupo Bimbo Company

Mexican multinational bakery product manufacturing company

Currently, daily inventory calculations are performed by sale's personal experiences.

Grupo Bimbo product

Question

1. Data Cleaning

2. Demand Mapping

3. Returned Factor

4. Demand Forecasting

Q1 Data Cleaning

- What we found in the dataset
- How we clean
    - Product grouping
    - Check wrong data

Q1: What we found in the dataset

- No identified product

- Not bread product

ProductSmp.csv

Producto_	NombreProducto
0	NO IDENTIFICADO 0
9	Capuccino Moka 750g NES 9
41	Bimbollos Ext sAjonjoli 6p 480g BIM 41
53	Burritos Sincro 170g CU LON 53
72	Div Tira Mini Doradita 4p 45g TR 72

- Same state but different name

town_state.csv

Agencia_ID	Town	State
1210	2059 Queretaro Balvanera	Queretaro de Arteaga
1232	2059 Queretaro Balvanera	Queretaro de Arteaga
2059	2059 QUERETARO BALVANERA	QUERETARO
2090	2090 AG. TEPEJI DEL RIO	QUERETARO
20599	2059 QUERETARO BALVANERA	QUERETARO
21719	2171 QUERETARO SAN PABLO	QUERETARO
21739	2173 SAN JUAN DEL RIO	QUERETARO
22090	2090 AG. TEPEJI DEL RIO	QUERETARO

Q1: What we found in the dataset

ClientSmp.csv

- Duplicate client

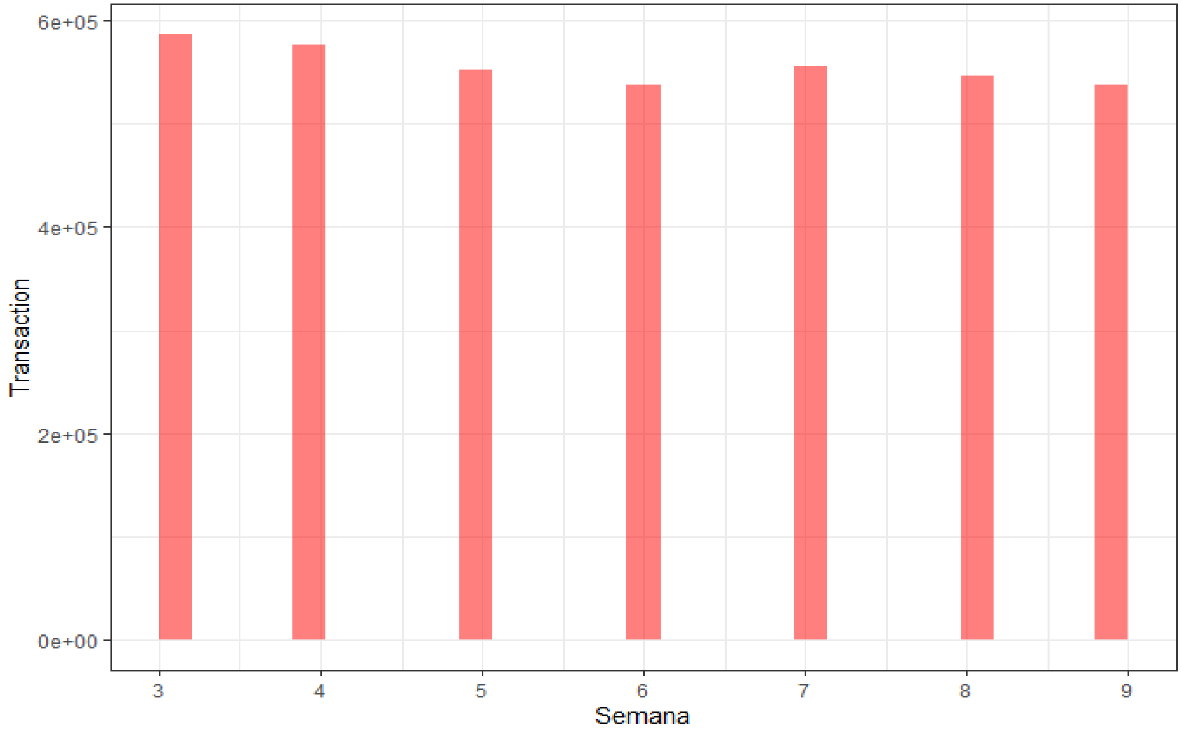
Cliente_ID	NombreCliente
0	SIN NOMBRE
1	OXXO XINANTECATL
2	SIN NOMBRE
3	EL MORENO
4	SDN SER DE ALIM CUERPO SA CIA DE INT
4	SDN SER DE ALIM CUERPO SA CIA DE INT

- No identified client

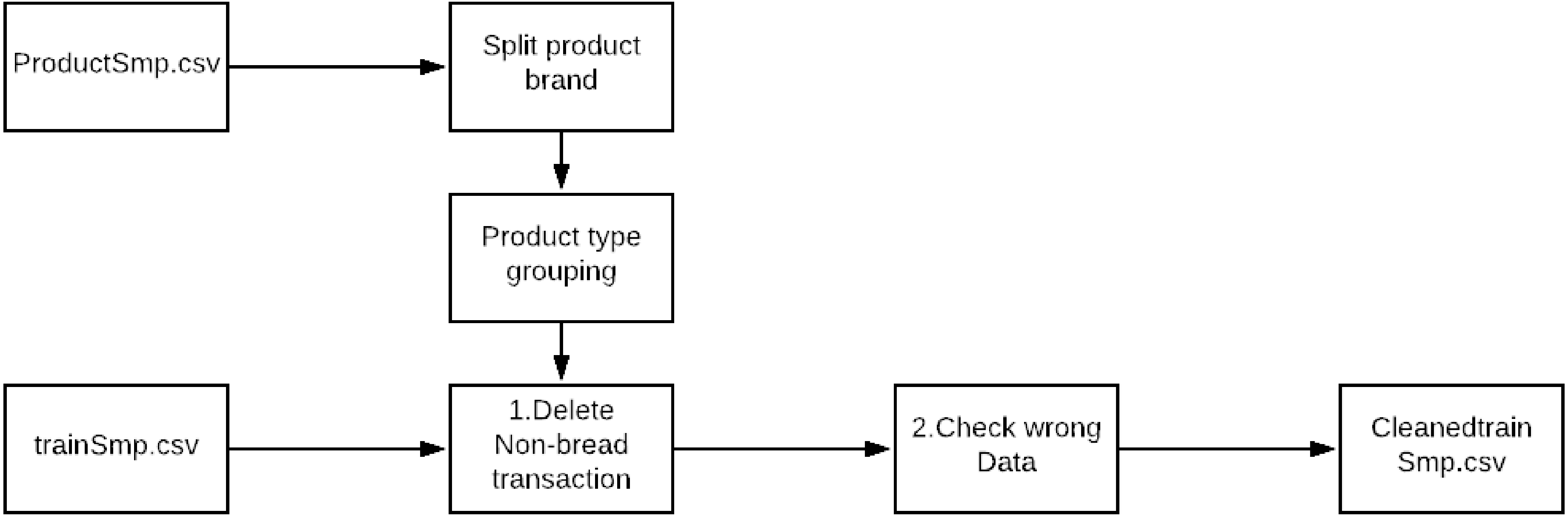
Cliente_ID	NombreCliente
4353446	NO IDENTIFICADO
4353447	NO IDENTIFICADO
4353448	NO IDENTIFICADO
4353449	NO IDENTIFICADO
4353450	NO IDENTIFICADO
4353452	NO IDENTIFICADO
4353453	NO IDENTIFICADO
4353454	NO IDENTIFICADO
4353460	NO IDENTIFICADO
4353461	NO IDENTIFICADO
4353465	NO IDENTIFICADO

Q1: What we found in the dataset

trainSmp.csv

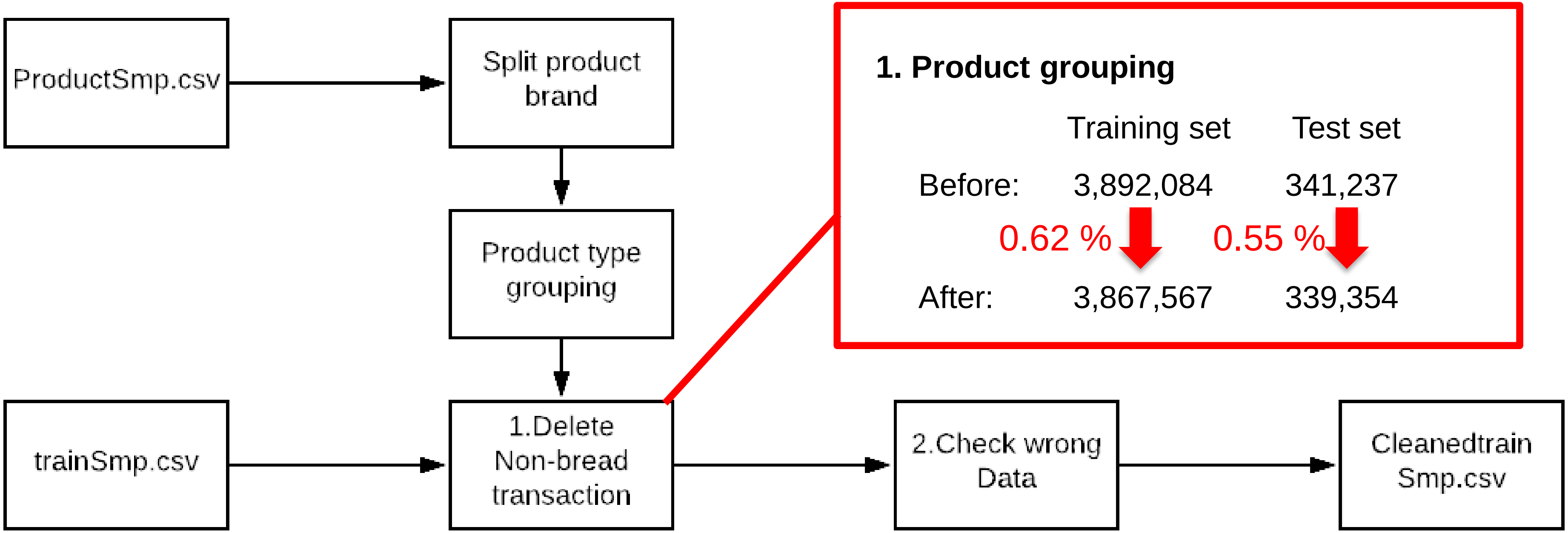


Q1: How we clean



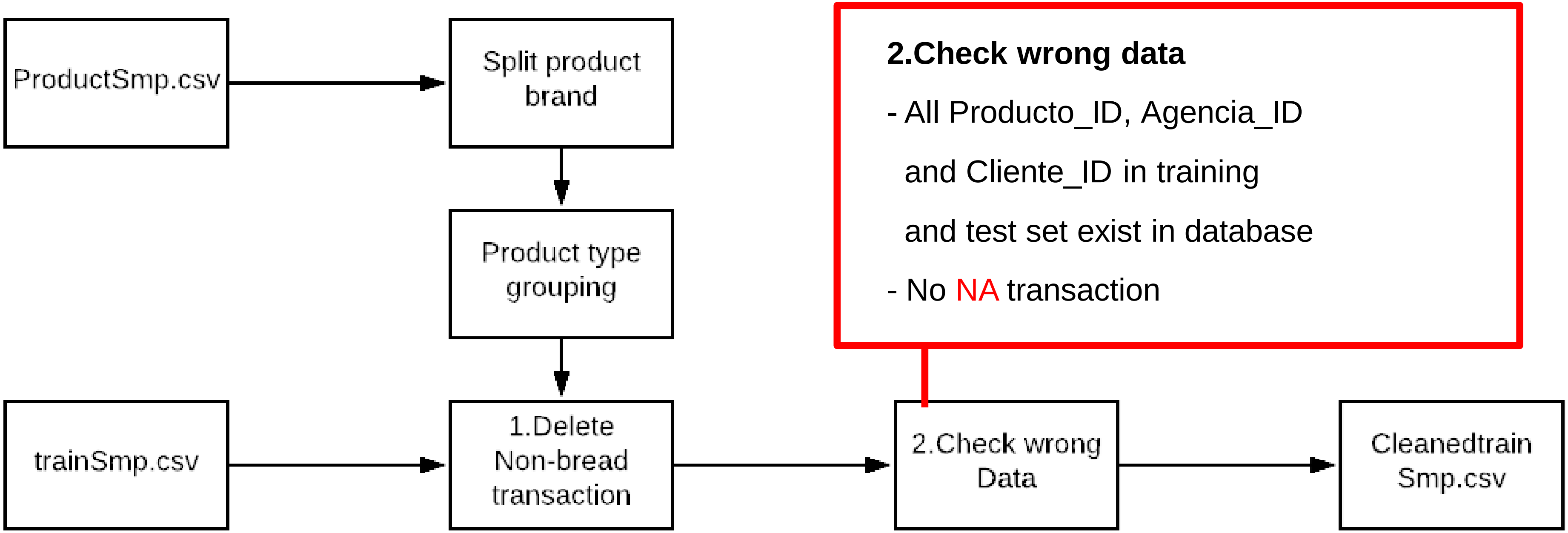
Data cleaning flow chart

Q1: How we clean



Data cleaning flow chart

Q1: How we clean



Data cleaning flow chart

Q1: Product Grouping

Group 1. Non-Bread brand

Group 3. Can’t Identified brand

Ex. 1 Kg , NO. IDENTIFICATION

Group 2. Mixing bread product brand

Q1: Product Grouping

Categorize all products in productSmp.csv

- Word frequency
- Manually examination

word	n
Pan	219
Prom	196
Fresa	148
Duo	101
Tubo	100
Vainilla	94
Deliciosas	92
Blanco	91
Chocolate	90
Tira	82
Gansito	80
Cj	77
Suavicremas	69
Galleta	66
Nuez	65
Multigrano	63
Pina	61
Barritas	59
Tortilla	59
Tostada	59
Frut	58
Bran	57
Principe	55
Mantecadas	53
Mini	52
Triki	51
Tostado	50
Integral	48
Roles	48
Bollos	46
Bimbollos	45
Tortillinas	45

Product name

Product type

Group 2. Mixing bread product brand

Q1: Product Grouping

Categorize all products in productSmp.csv

Group 2. Mixing bread product brand

19 Product groups

Pie	91
Pizza	5
Roll	34
Short bread	38
Snack	233
Tortilla	150
White bread	146
Whole grain	184

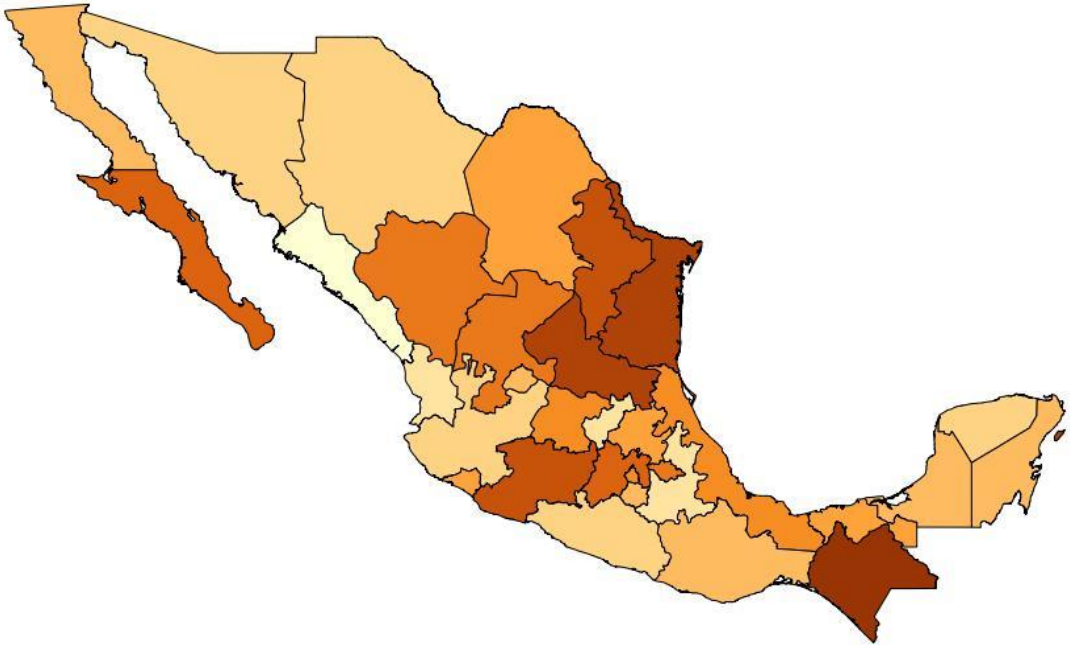
Bigote	22
Biscuit and cracker	94
Bread crumbs	25
Bun	139
Cake	180
Cookie	666
Croissant	16
Donut	41
Hot dog	82
Muffin	7
Nacho	143

Q1: Cleaned training set

	Semana	Agencia_ID	Canal_ID	Ruta_SAK	Cliente_ID	Producto_ID	Demanda_uni_equil
1	3	1110	7	3301	818913	1146	38
2	3	1110	7	3301	818913	31688	8
3	3	1110	7	3301	4328697	1146	10
4	3	1110	7	3302	319684	1250	2
5	3	1110	7	3302	1298872	1240	4
6	3	1110	7	3302	1298872	1250	5
3867563	9	25759	1	5517	4357997	35635	4
3867564	9	25759	1	5517	4388275	35132	3
3867565	9	25759	1	5517	4388280	32861	4
3867566	9	25759	1	5517	4488100	37401	1
3867567	9	25759	1	5517	4494351	37024	1

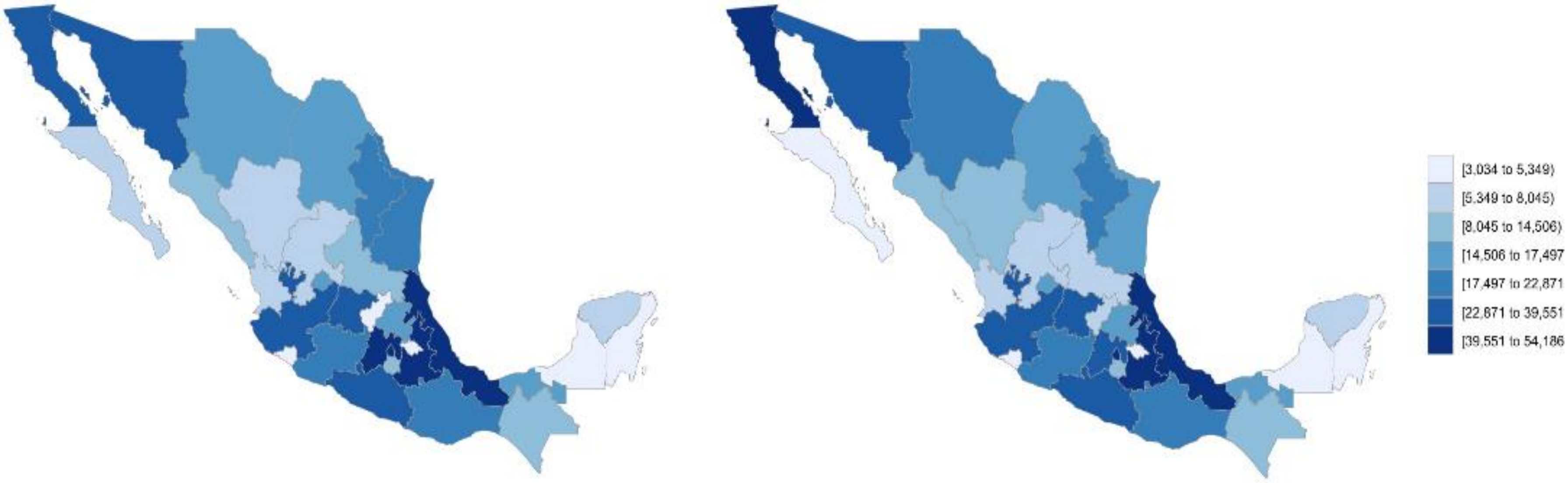
Q2 Demand Mapping

- How we generate map
- Demand map of tortilla product
- What insights do we gain



Q2: How we generate map

Generate mexico map using mxmaps package

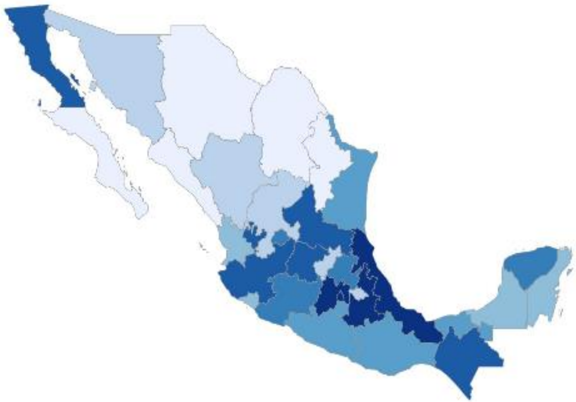


week 3

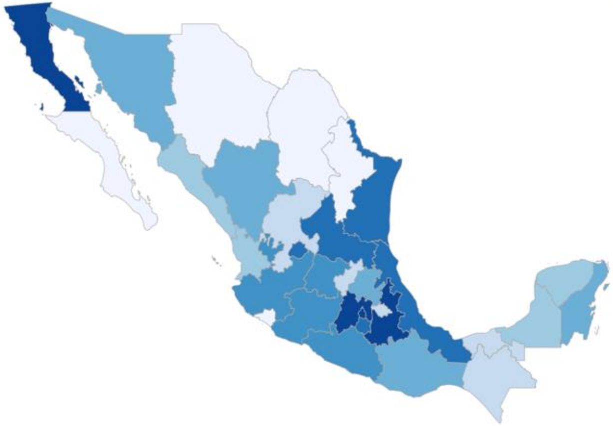
week 9

Example: Demand map of cookie product

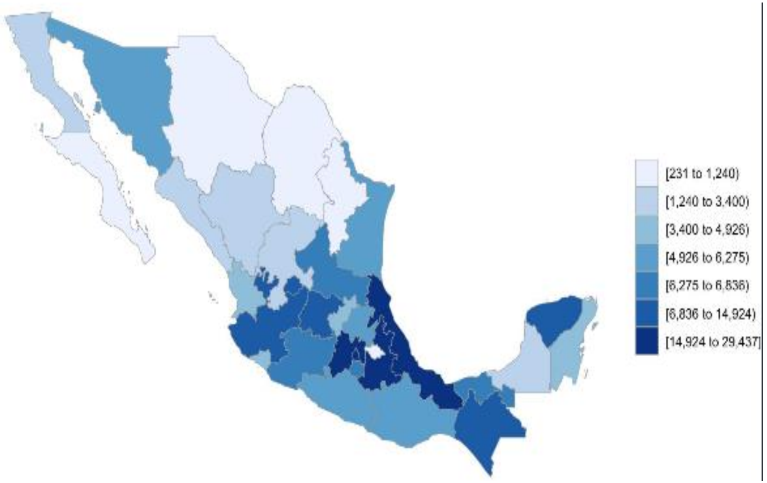
Q2: Demand map of tortilla product



week 3



week 6



week 9

Q2: Demand map of tortilla product

week 3

week 6

week 9

Q2: Why tortilla demand deceasing in the North of Mexico?

What is Tortilla?

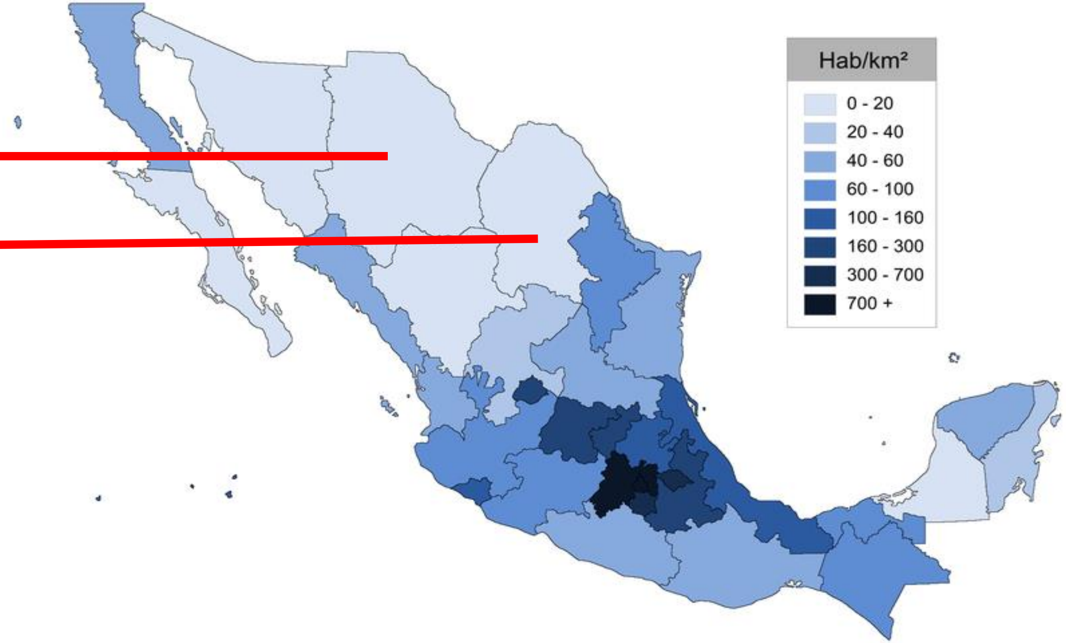
- it's a type of thin flatbread, typically made from corn or wheat.
- commonly used in burritos, tacos, fajitas, and other Tex-Mex foods.

Q2: Why tortilla demand deceasing in the North of Mexico?

Population in Mexico 2010

Chihuahua: 3,406,465 (2.98%)

Coahuila: 2,748,391 (2.41%)



North Mexico is a plateau region

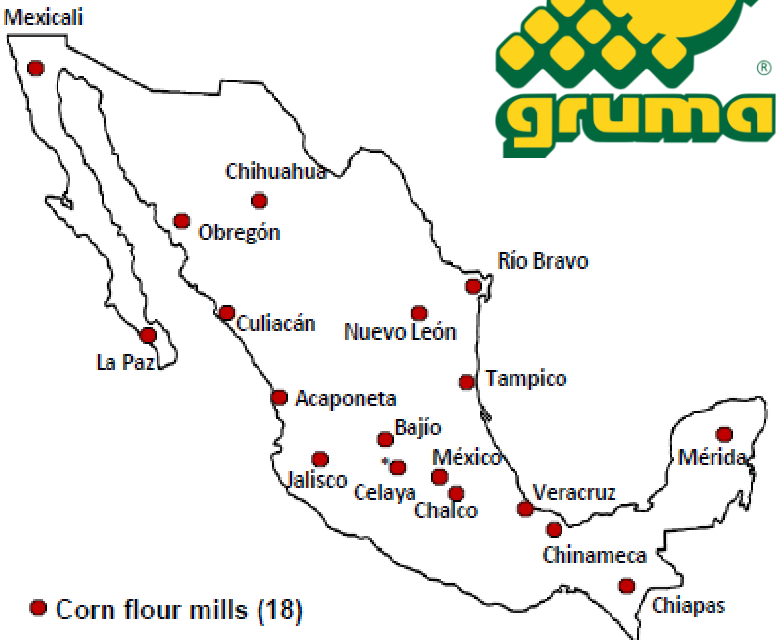
Q2: Why tortilla demand deceasing in the North of Mexico?

Major Tortilla Competitor: GRUMA Company

Gimsa Company

- ❖ Mexico's leading corn & wheat flour manufacturer
- ❖ 1st Market share of corn flour in Mexico approximately 74%

GIMSA
GRUPO INDUSTRIAL MASECA S.A.B DE C.V.



Source: Mexico And Mexico City In The World Economy

Q2: Why tortilla demand deceasing in the North of Mexico?

Major Tortilla Competitor: GRUMA Company

Mission Mexico Company

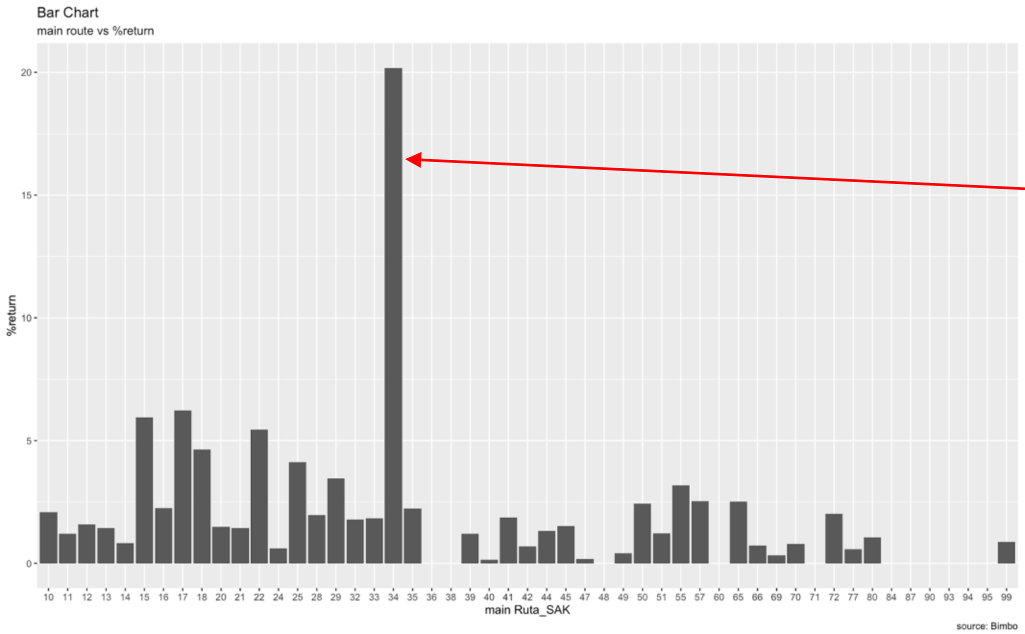
- It is a producer of corn flour tortillas, wheat flour tortillas, and tortilla chips
- Two plants and distribution centers located principally in Northern Mexico

Q3 Returned Factor

- Our scope
- Statistical approach
- Visualization approach

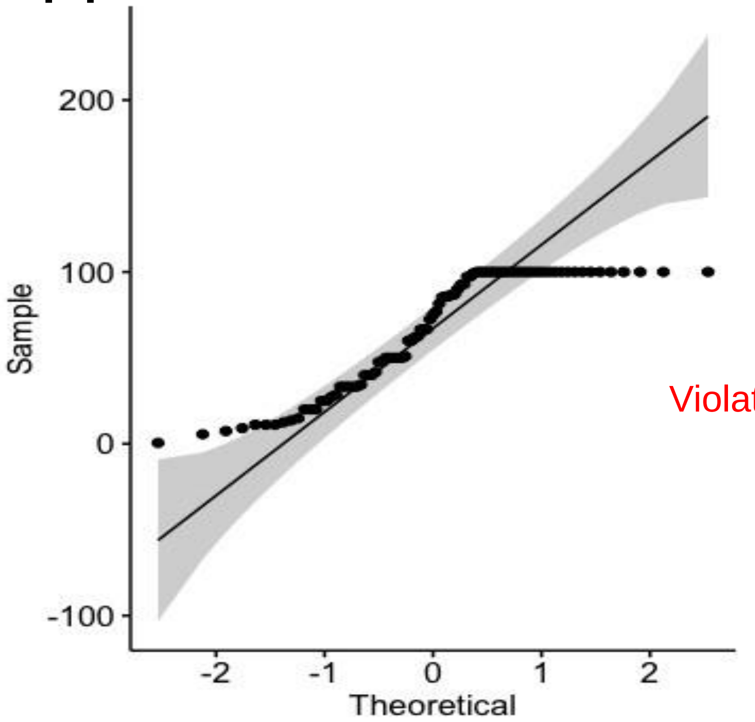
Q3: Our scope: High returns road

Focus on highest returns road (route) 34XX




```
> route_2digit[order(-route_2digit$percentRT),]
# A tibble: 51 x 4
   Ruta_2digit Dev_uni_proxima Venta_uni_hoy percentRT
   <chr>                 <int>         <int>     <dbl>
 1 34                    44616        176457     20.2
 2 17                       90          1355      6.23
 3 15                     4508         71348      5.94
 4 22                     2064         35832      5.45
 5 18                      444          9148      4.63
 6 25                      718         16738      4.11
 7 29                     2628         73181      3.47
 8 55                      502         15338      3.17
 9 57                     5573        215208      2.52
10 65                      248          9620      2.51
# … with 41 more rows
```

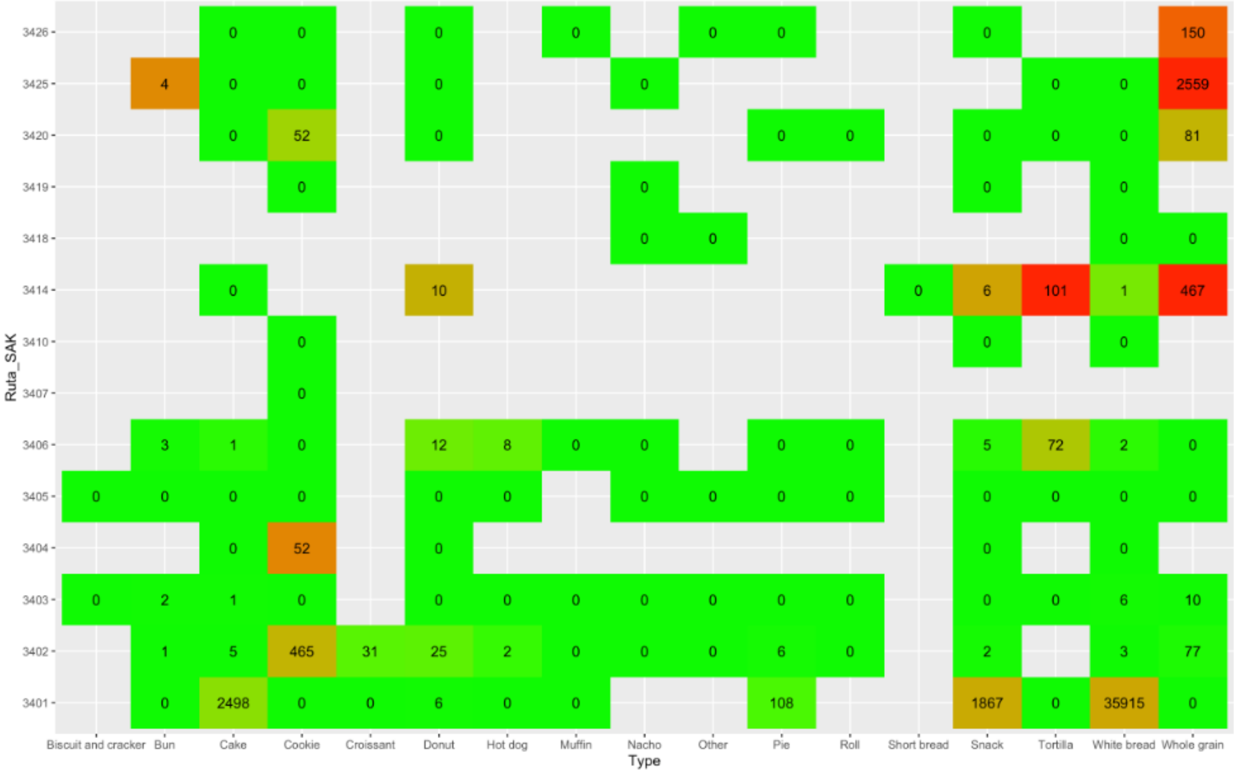
Q3: Statistical Approach



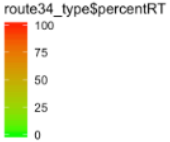
Normal plot of return rate in route 34

Violate ANOVA assumptions

Q3: Visualization Approach: Return rate by Product type



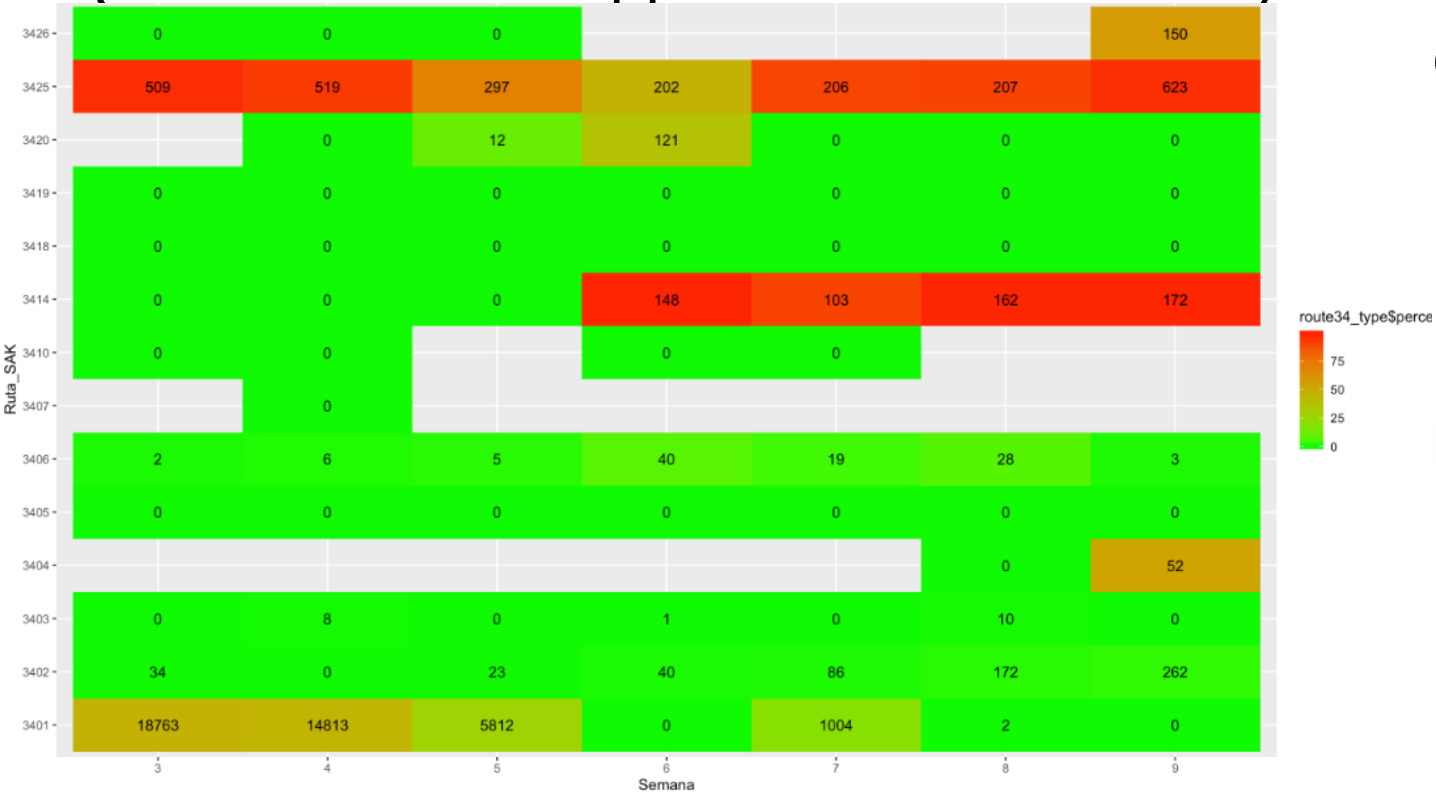
Group.1	percentRT
Whole grain	32.827190
Tortilla	19.230769
Cookie	11.820695



Average Return rate Of all route 34xx

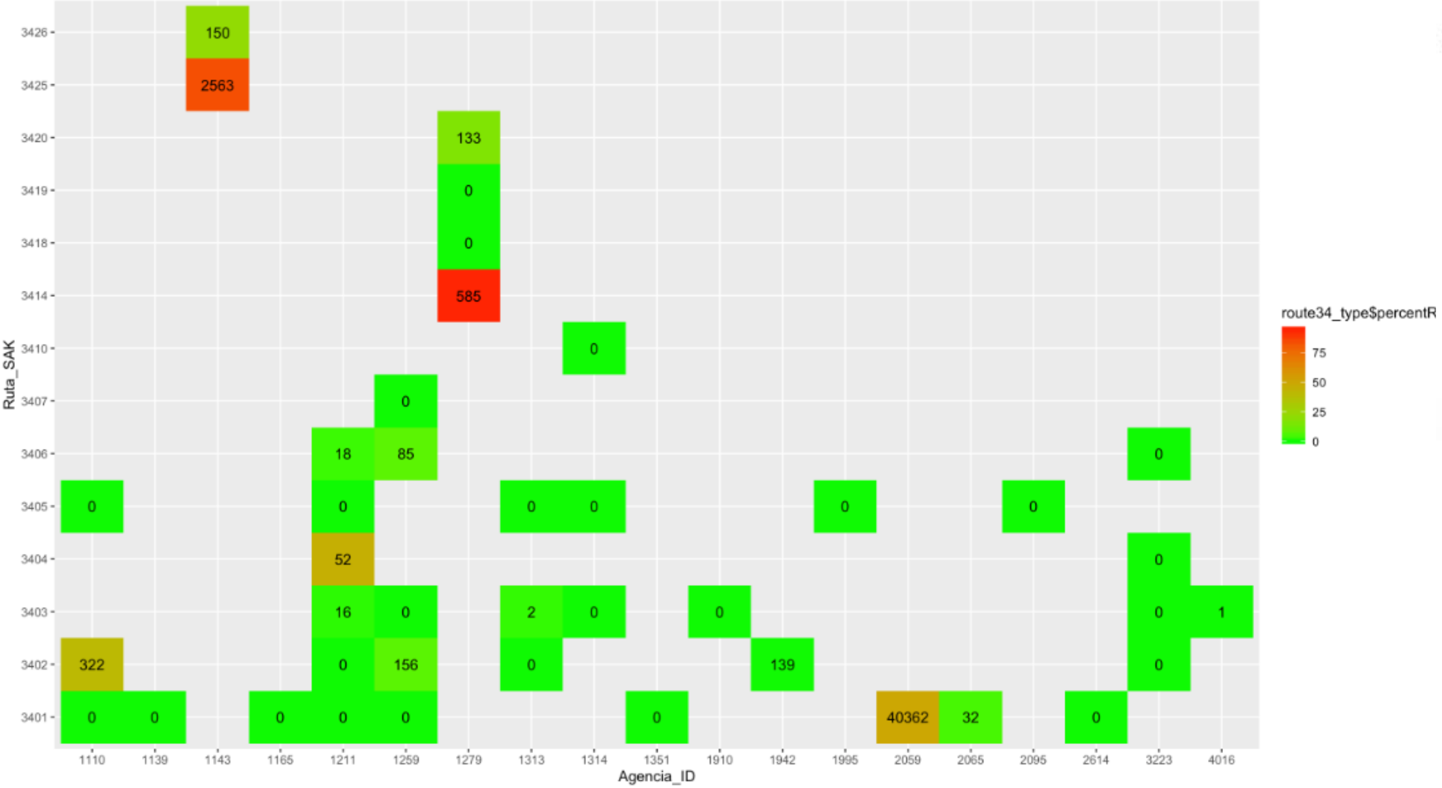
Criteria: If average return Rate > 10%, we conclude That It affects the return

Q3: Visualization Approach: Return rate by Week



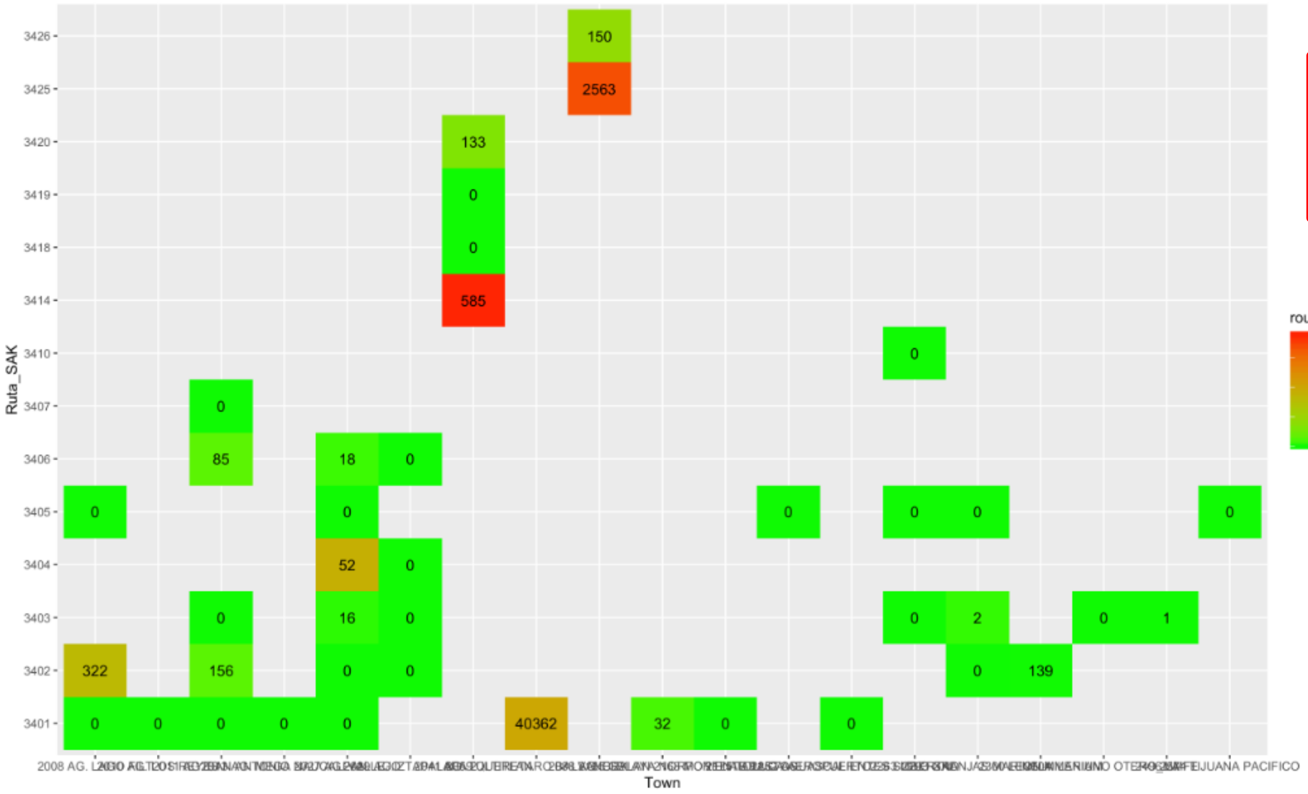
Group.1	percentRT
9	25.67519
7	18.72733
8	18.03054
6	17.34739
3	13.01191
4	10.77261
5	10.10691

Q3: Visualization Approach: Return rate by Agent_ID

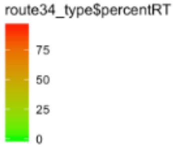


Group.1	percentRT
1143	52.825720
2059	49.559196
1279	27.704607
1110	13.467169
1211	8.438536
2065	4.551920

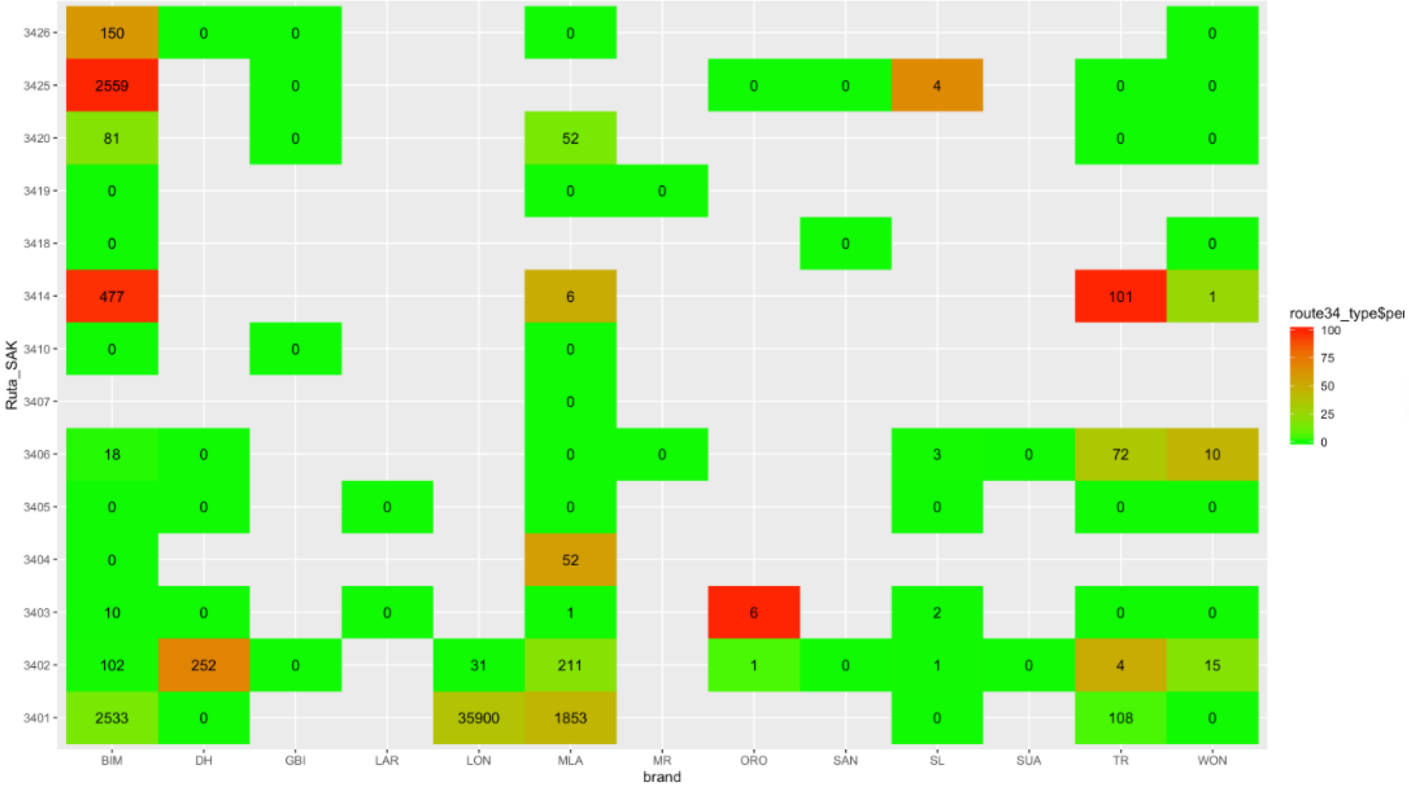
O3: Visualization Approach: Return rate by Town



	Group.1	percentRT
2088	AG. CEYLAN	52.825720
2059	QUERETARO BALVANERA	49.559196
2041	AG. TULTITLAN	27.704607
2008	AG. LAGO FILT	13.467169
2027	AG. VALLEJO	8.438536
2155	CELAYA NORPONIENTE	4.551920

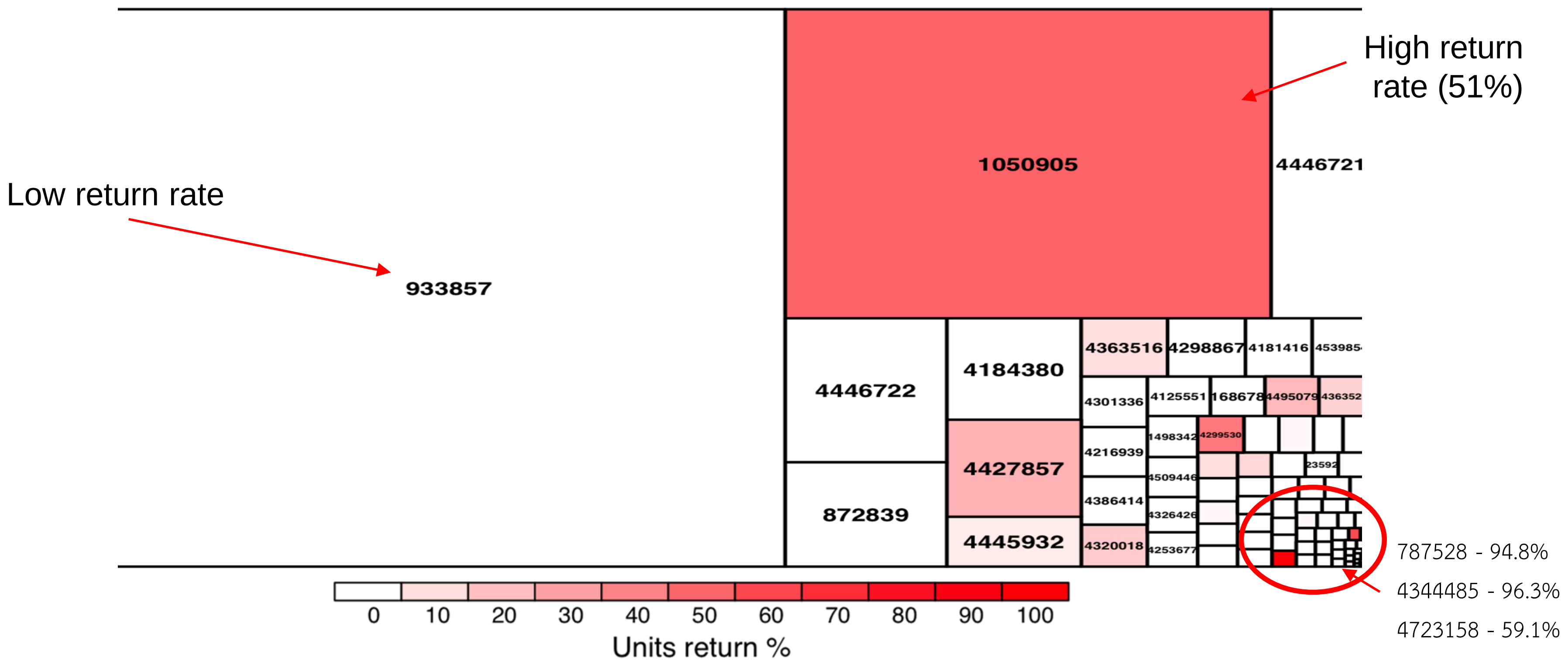


Q3: Visualization Approach: Return rate by Brand



Group.1	percentRT
ORO	35.00000
TR	23.69140
BIM	22.22756
LON	18.50855
MLA	15.42392
DH	11.57025

Q3: Visualization Approach: Return rate by Client, Sales unit



Q3: Summary of returned factors

Significant returned factors in route 34XX

- ❖ Agent_ID
  - ❖ 1143,2059,1279,1110
- ❖ Week
- ❖ Town_ID
  - ❖ 2088,2059,2041,2008
- ❖ Type
  - ❖ Whole grain,Tortilla,Cookie and Bun
- ❖ Brand
  - ❖ ORO,TR,BIM,LON,MLA and DH

Q4 Demand Forecasting

- KPI: RMSLE
- Technique: XGBoost
- Feature engineering
- Result

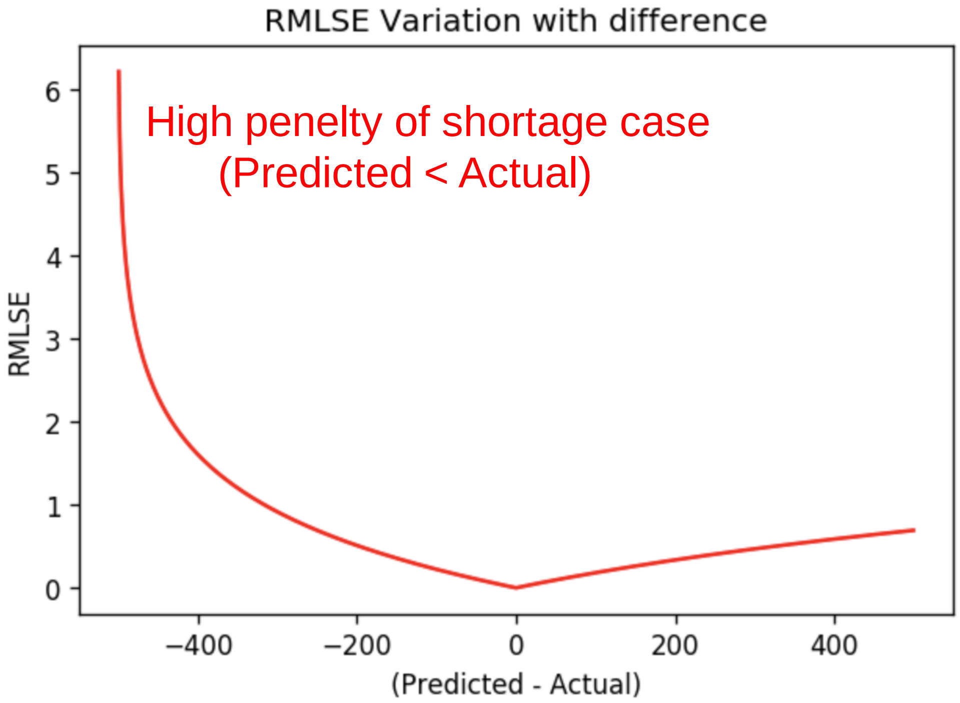
Q4 KPI: RMSLE (Root mean square log error)

RMSLE Formulae

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(x_i+1) - \log(y_i+1))^2}$$

Actual Value

Predicted Value

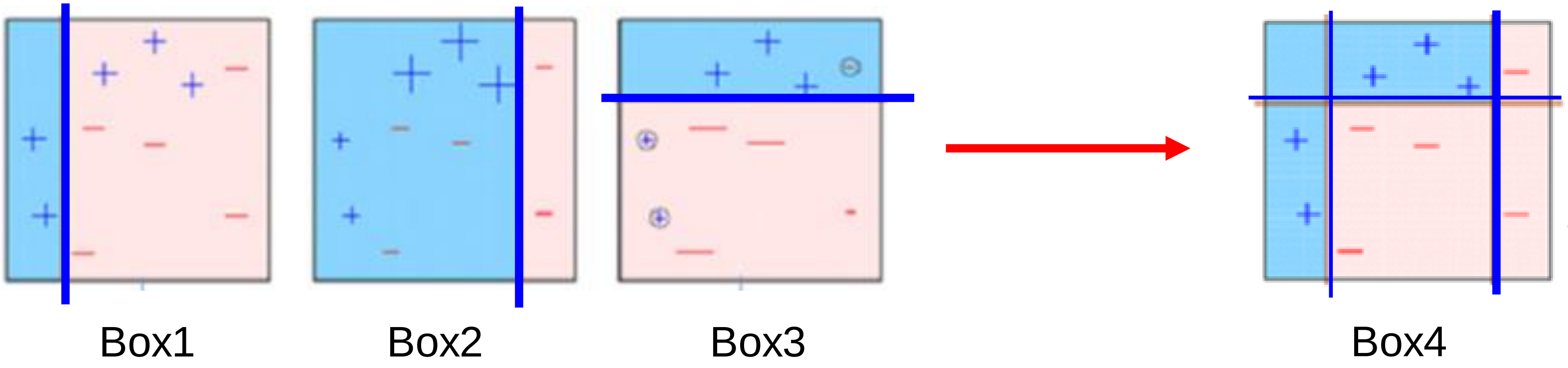


Q4: XGBoost (Extreme gradient boosting)

- Efficient version of gradient boosting
- Combination of linear model solver and tree learning algorithm

Concept of Boosting...convert weak learners into strong learners

Trees are grown using the information from a previously grown tree



Q4: XGBoost (Extreme gradient boosting)

- Efficient version of gradient boosting
- Combination of linear model solver and tree learning algorithm

Why is it good?

- **Regularization:** avoid overfitting in linear and tree-based models
- **Parallel Computing:** It is enabled with parallel processing, by default, it would use all the cores of your laptop/machine.
- **Flexibility:** it can handle with regression, classification, and ranking problems,
- **Efficient handling of missing data**

Q4: Feature Engineering

Extract new feature: Mean of demand by Product ID and Client ID

```
> demand_mean_client
  Producto_ID Cliente_ID  Mean_byPC
1          72        772  4.0000000
2        1109        772  1.0000000
3        1146        772  2.0000000
4        1240        772  6.0000000
5        1250        772  4.0000000
6        1278        772 13.6666667
7        1284        772 11.8333333
```

Q4: Result

Training set (Week 3-7)

Train

XGBoost Predictor

Test

Validation set (Week 8-9)

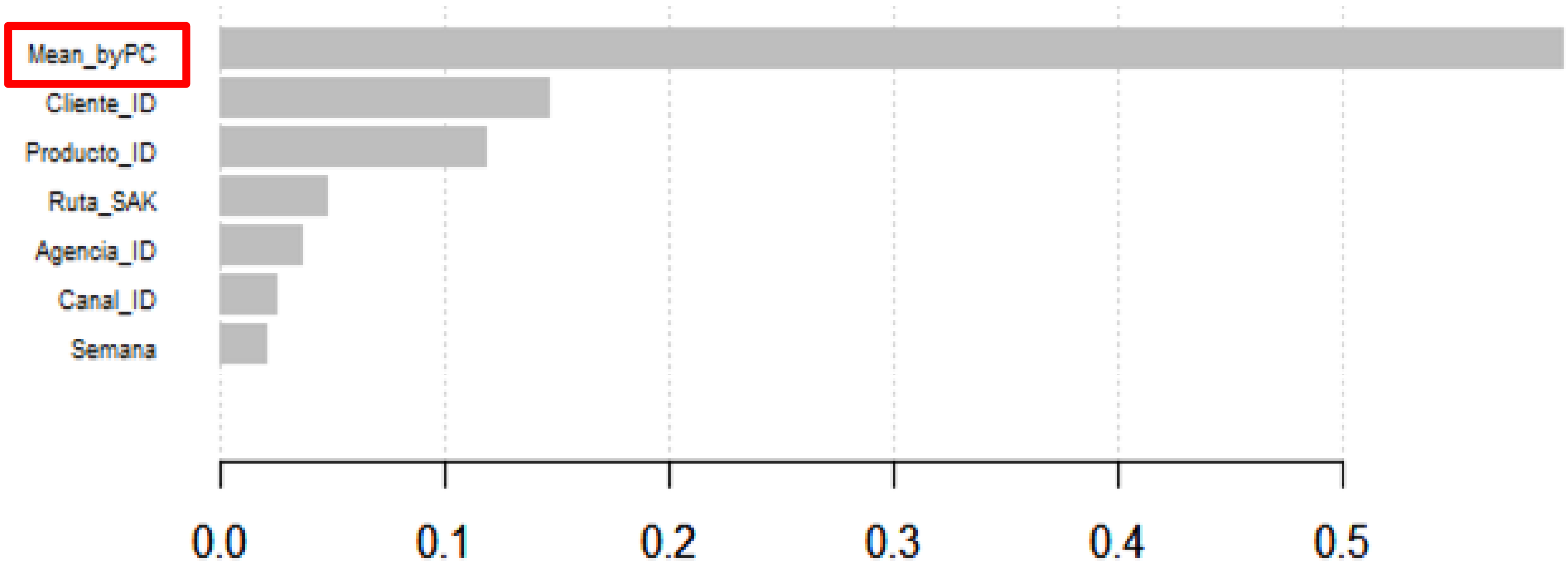
Hyperparameters

- Set.seed(20)
- eta (learning rate) = 0.1
- booster = "gbtree”
- objective = "reg:linear”
- max_depth = 10

Result

```
> ## cal rmsle
> rmsle(ts_label, xgbpred)
[1] 0.5187931
```

Q4: Result – Feature Importance



Summary

1. Data Cleaning

- ✓ Product grouping
- ✓ Checking wrong data

3. Returned Factor

- ✓ Insight Highest return Route (34XX)

2. Demand Mapping

- ✓ mxmaps package
- ✓ Tortilla insights
  - Population relating
  - Competitor

4. Demand Forecasting

- ✓ Technique: XGBoost
- ✓ Result: RMSLE = 0.518

CHULA ΣNGINEERING

Foundation toward Innovation