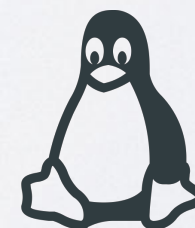
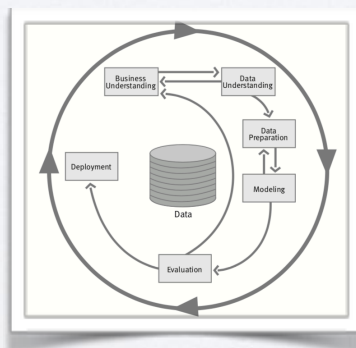


# WALMART RECRUITING:TRIP TYPE CLASSIFICATION

Itzel Muñoz | 22803  
Héctor Corro | 14350  
Uriel Miranda | 77508



# BUSINESS UNDERSTANDING



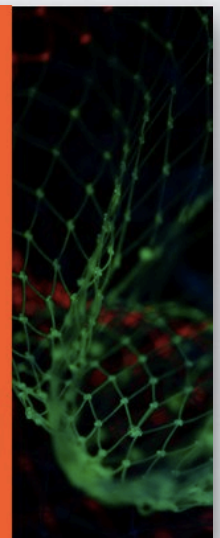
- 20,000 Tiendas
- 28 Países
- THE DATA CAFE
- 2.5 PB / Hr

“

A **Petabyte** is the equivalent of **4,000** digital photos per day, over your entire life.

Lifewire.com

”



# DATA UNDERSTANDING



```
graph TD; TripType --> VisitNumber; VisitNumber --> Weekday; Weekday --> Upc
```

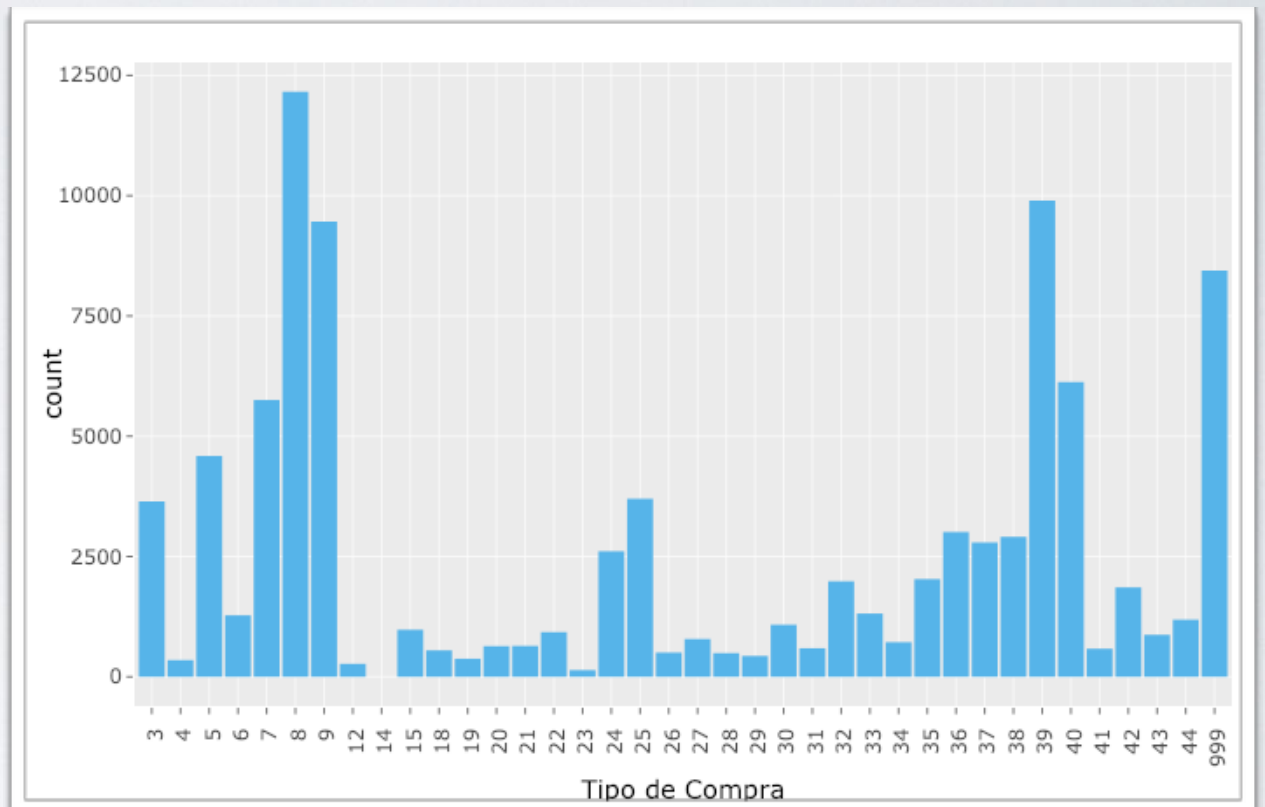
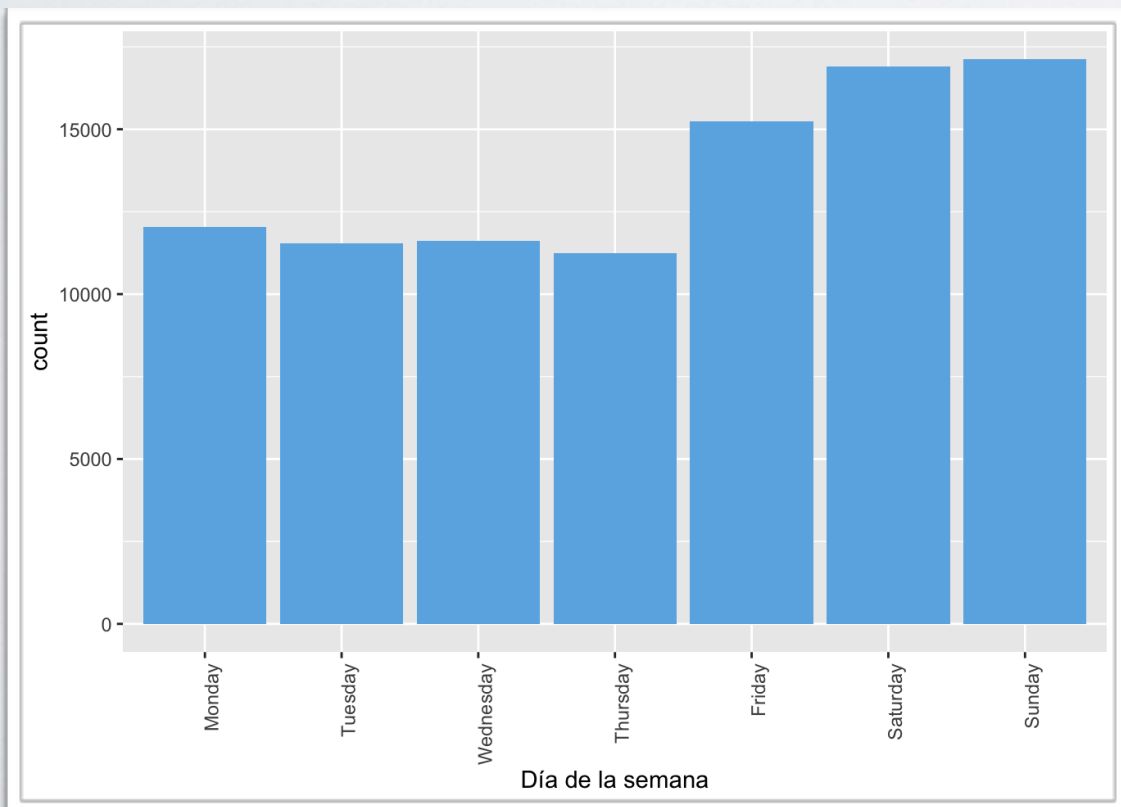
TripType  
~~VisitNumber~~  
~~Weekday~~  
Upc

# Weekday

U<sub>pc</sub>

ScanCount	DepartmentDescription	FinelineNumber
-----------	-----------------------	----------------

FinelineNumber

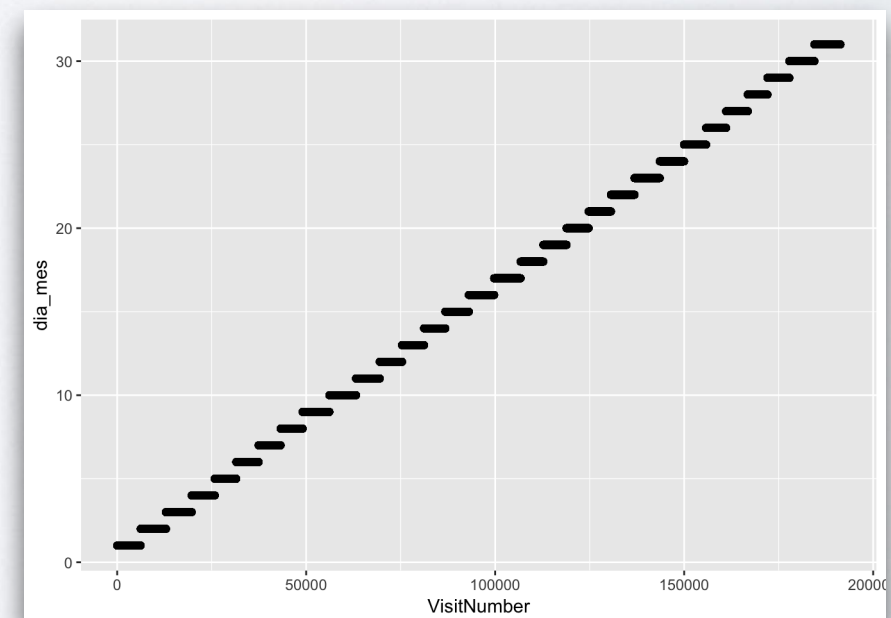
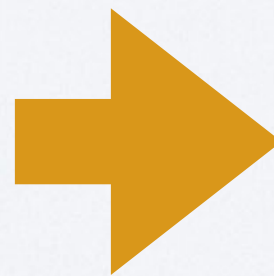
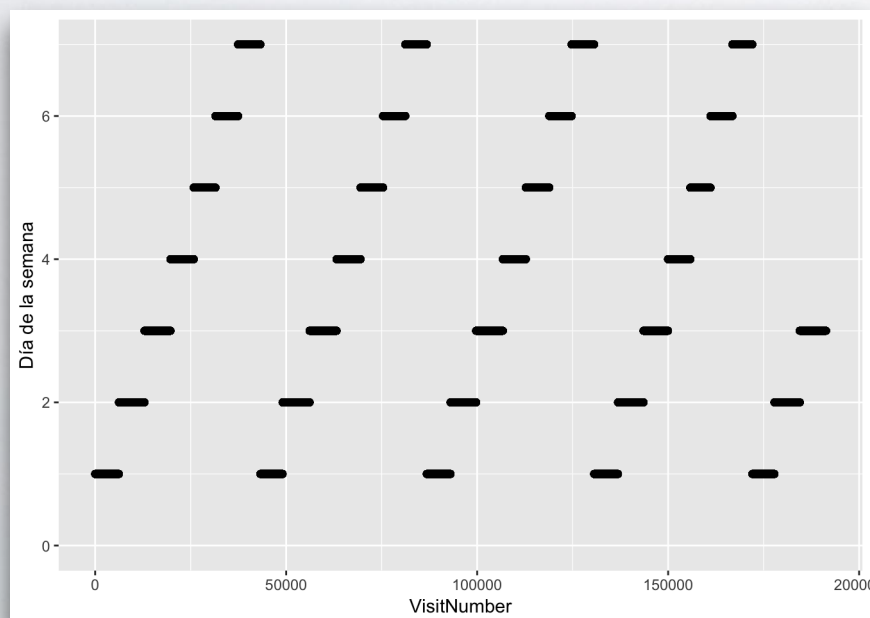
[illegible]



# DATA PREPARATION

## 6 Variables explicativas

- Total de productos comprados en la visita.
- Indicadora de devolución de productos.
- Total de productos devueltos en la visita.
- Total de departamentos visitados, si se compró al menos un producto en el departamento éste cuenta cómo 1a en la suma de departamentos visitados.
- Productos diferentes comprados respecto a UPC, es decir cuántos productos con diferente código UPC se compraron en la visita
- Día del mes en que se realizó la visita, se explica más adelante esta variable.



# FEATURE ENGINEERING

6 Variables explicativas  
647,054 Filas



2 Variables explicativas  
eliminadas



71 Variables explicativas  
~95,000 Filas

# MODELS

■ ■ ■



Modelo	Parámetros	Tiempo Ajuste	Public Score
Regresión Multinomial	penalización L1, C = 1	3 hrs	1.3185
Bosque Aleatorio	depth = 100, feat = sqrt, split = 10, n_est = 1500	4 hrs	2.1320
Máquina de Soporte Vectorial	kernel lineal, C = 10	5 hrs	1.0820
Red Neuronal	alpha 100, 30 capas, 1000 iter, semilla 0	4 hrs	1.0370

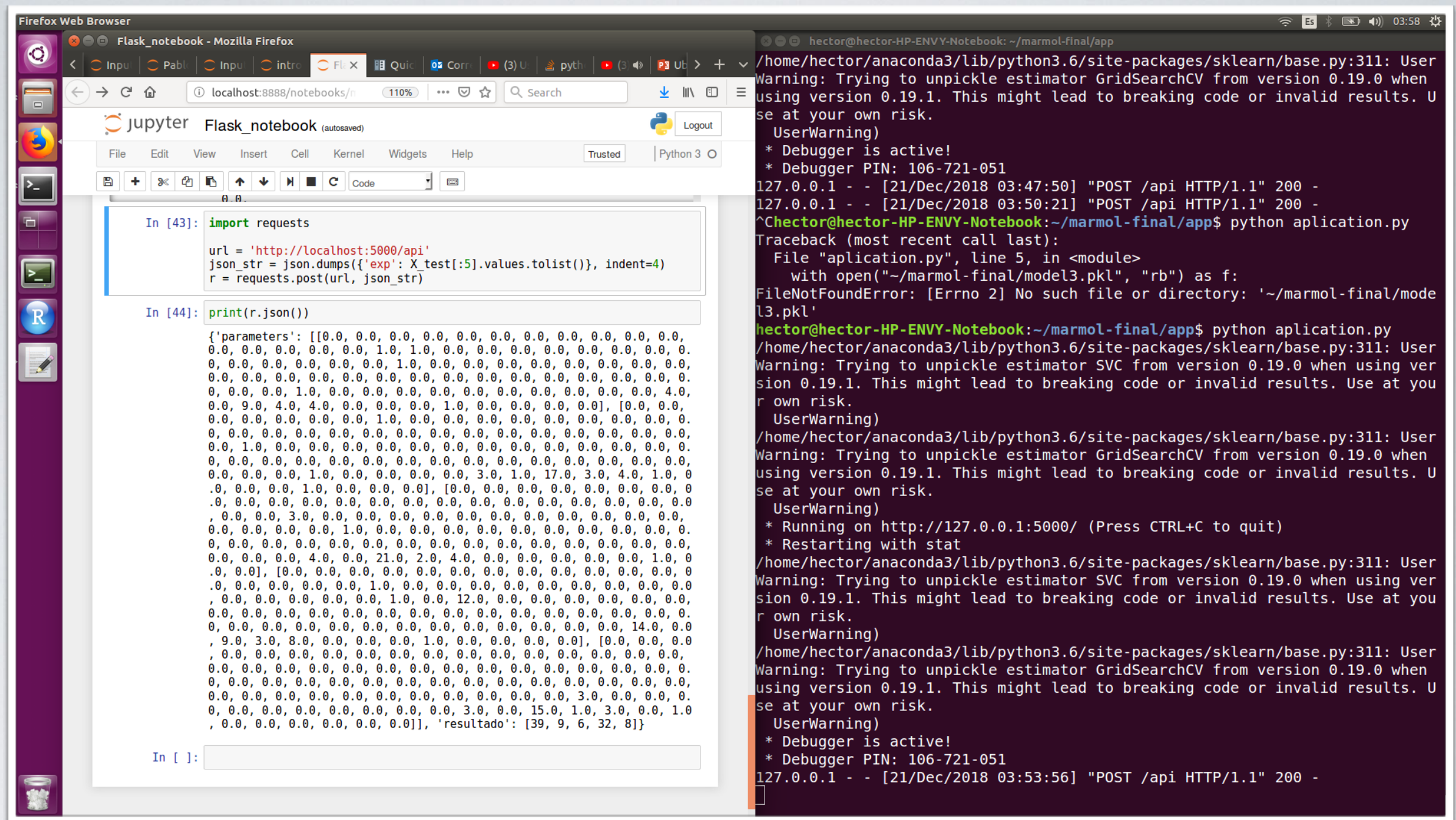


# EVALUATION

Lugar 445 de 1047  
Public Score 1.02535  
Percentil 42%

XBoost (Out of the Box)  
Public Score 0.94165  
Lugar 441 de 1047  
Percentil 39%

Overview	Data	Kernels	Discussion	Leaderboard	Rules	Team	My Submissions	Late Submission
-m Message								?
8 submissions for Itzel Fernández de Córdova								Sort by Most recent
All Successful Selected								
Submission and Description				Private Score	Public Score	Use for Final Score		
<b>pred_svm.csv</b> 9 hours ago by Itzel Fernández de Córdova modelo_svm				1.06870	1.08196	<input type="checkbox"/>		
<b>pred_reg.csv</b> 12 hours ago by Itzel Fernández de Córdova modelo_reg				1.30855	1.31850	<input type="checkbox"/>		
<b>pred_mas_rf.csv</b> 20 hours ago by Itzel Fernández de Córdova modelo_rf				2.11070	2.13202	<input type="checkbox"/>		
<b>pred_nn2.csv</b> 21 hours ago by Itzel Fernández de Córdova modelo_alpha10				1.06388	1.07411	<input type="checkbox"/>		
<b>pred_om.csv</b> 21 hours ago by Itzel Fernández de Córdova modelo_om				0.94165	0.96473	<input type="checkbox"/>		
<b>pred_nn1.csv</b> 21 hours ago by Itzel Fernández de Córdova modelo_alpha100				1.02535	1.03705	<input type="checkbox"/>		



# DEPLOYMENT

## Flask