BARF – Selling Quality Dog Food in the US

Assignment
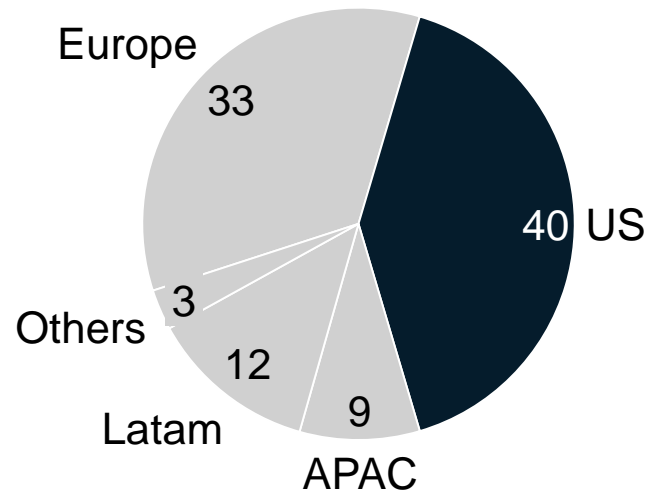
Jonas Debrabandere, Margret Dupslaff, Felipe Hollanda Lima, Sebastian Gili, Talha Reza, Bruh Terfie

30.05.2020

US pet food market is a $30B industry, of which dog food is 67%

Global pet food market (2016),
% (100 = $75.2B)

Europe
33

Others
3

Latam
12

APAC
9

40 US

US pet food market
% 100% = $30 B

Dog food   67%

Cat food   29%

Others[1]   4%

2016

Problem statement

Status Quo	Challenges	Solution
- Most Pet allergies caused by commercial dog food - Before commercial pet food, dogs barely experienced chronic allergies, ear infections, hot spots, itchy skin "BARF" (Biologically Appropriate Raw Food) → eradicates allergies	- Time-consuming to prepare and expensive compared to kibbles - Some breeds of dogs are dramatically affected by allergies compared to others - Seba Gili's wife to the rescue – want to set up a business to provide this	A model that could allow the team to predict where these dogs live, based on their family's general information, could provide the solution!

Input data provides total number of allergic dogs in New York
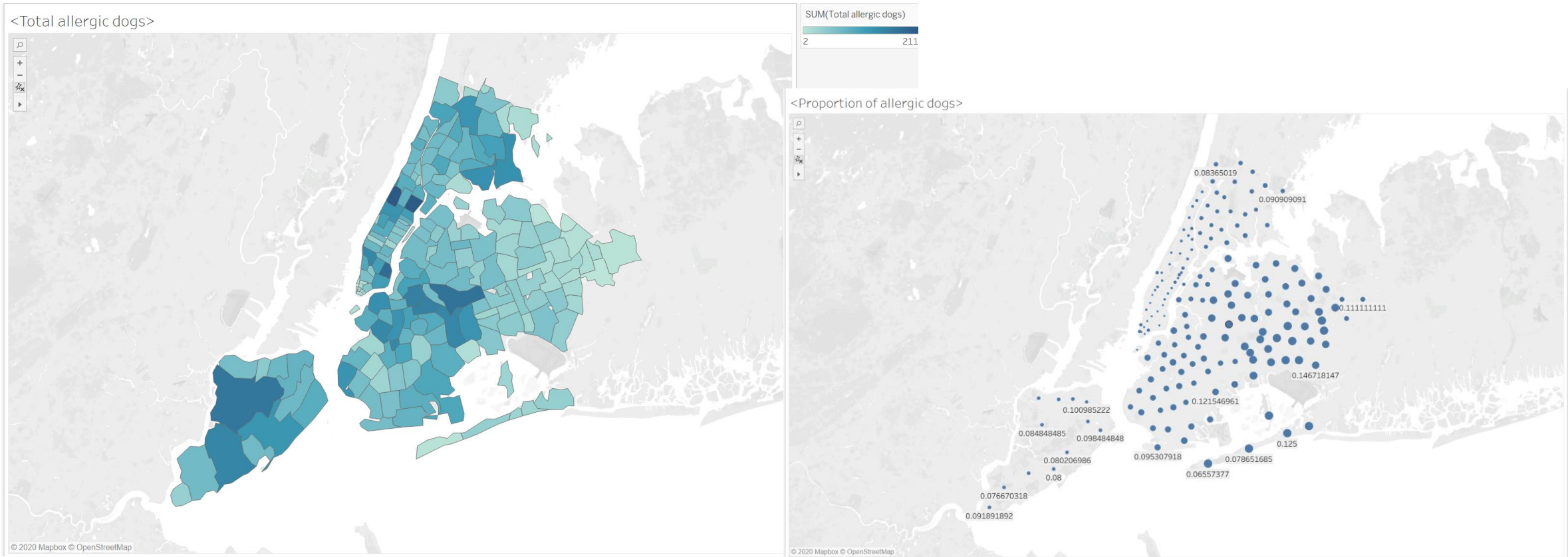
Input data from New York based on income level, population, ethnicity, type of work, household size, etc.
→ **Need to apply the data on Florida**

Implementation

Objective	Implementation	Standard Error of the model
1 **Total # dogs in Florida based on New York data** • Income • Demographics • Licenses of dogs • ZIP code	• Simple linear regression **1a** Simple linear regression using log of independent variables • Random Forrest • Random Forrest using log of independent variables • Principle Component Analysis (scaled) • Principle Component Analysis (not scaled)	• ~500 • ~439 • 512.35 • 530.23 • 521.60 • 518.81
2 **Percentage of allergic dogs in Florida**	**2a** Random Forest prediction of the % of allergic susceptible dogs using PCA	(see table below)

Reference / Prediction	0	1
0	8	1
1	2	24

Log-Linear regression

- Call:
- lm(formula = Total_dogs ~ Black + IncomePerCap + Poverty + Service +
-     Construction + Drive + Transit + OtherTransp + WorkAtHome +
-     Employed + PublicWork + Unemployment, data = linear_regression_log)

- Residuals:
-      Min   1Q   Median   3Q   Max
- -1.76710 -0.27290 0.03406 0.38544 1.62535

- Coefficients:
-              Estimate Std. Error t value Pr(>|t|)
- (Intercept) -25.60019   4.64349 -5.513 1.80e-07 ***
- Black   -0.16450   0.05104 -3.223 0.001601
- IncomePerCap 1.87315   0.32819   5.708 7.30e-08 ***
- Poverty   0.92172   0.19527   4.720 5.97e-06 ***
- Service   0.79170   0.21050   3.761 0.000254 ***
- Construction 0.28263   0.17100   1.653 0.100763
- Drive   0.25226   0.15582   1.619 0.107876
- Transit   1.04315   0.23918   4.361 2.59e-05 ***
- OtherTransp   0.22928   0.11757   1.950 0.053302 .
- WorkAtHome   0.29947   0.11779   2.542 0.012172 *
- Employed   0.24349   0.14171   1.718 0.088113 .
- PublicWork   0.45227   0.18041   2.507 0.013405 *
- Unemployment -0.27250   0.19835 -1.374 0.171840

- Signif. codes: 0 '***' 0.001 '' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.549 on 131 degrees of freedom
Multiple R-squared: 0.5425, Adjusted R-squared: 0.5006
F-statistic: 12.94 on 12 and 131 DF, p-value: < 2.2e-16

INSEAD

Confusion Matrix and Statistics

          Reference
Prediction 0 1
         0 8 1
         1 2 24

           Accuracy : 0.9143
             95% CI : (0.7694, 0.982)
No Information Rate : 0.7143
P-Value [Acc > NIR] : 0.004065

              Kappa : 0.7835

Mcnemar's Test P-Value : 1.000000

        Sensitivity : 0.8000
        Specificity : 0.9600
     Pos Pred Value : 0.8889
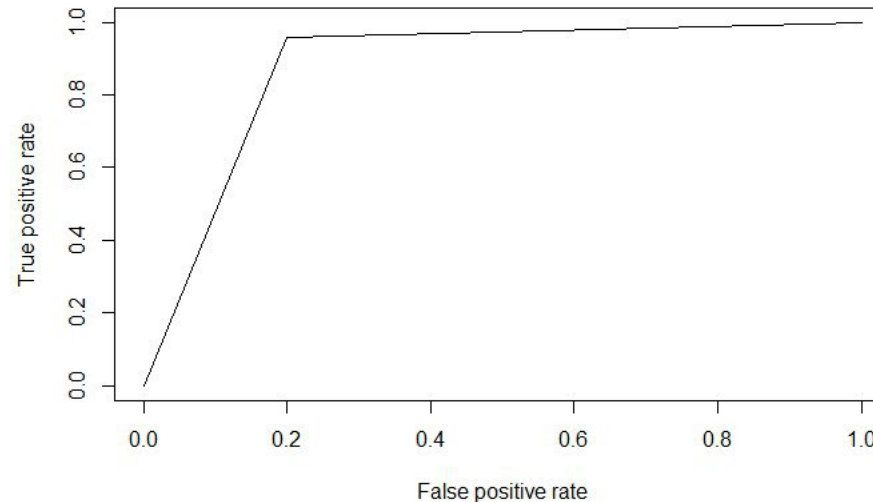   Neg Pred Value : 0.9231
         Prevalence : 0.2857
     Detection Rate : 0.2286
Detection Prevalence : 0.2571
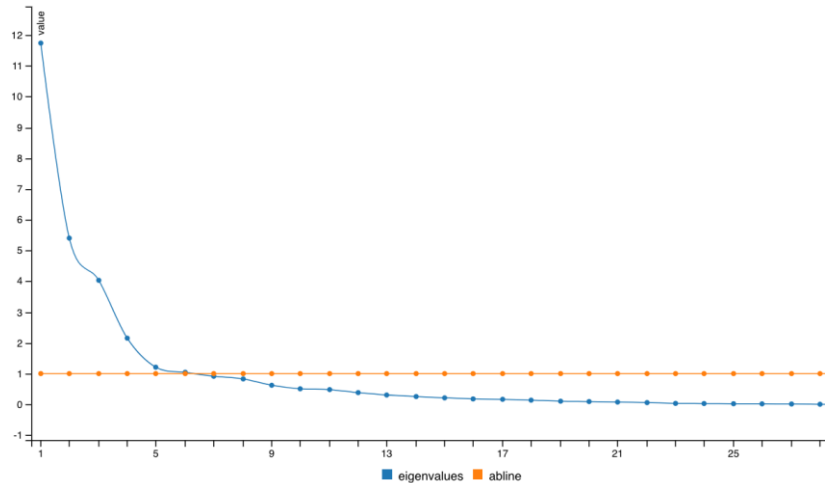   Balanced Accuracy : 0.8800

     'Positive' Class : 0

AUC = 0.88

Threshold = 11%

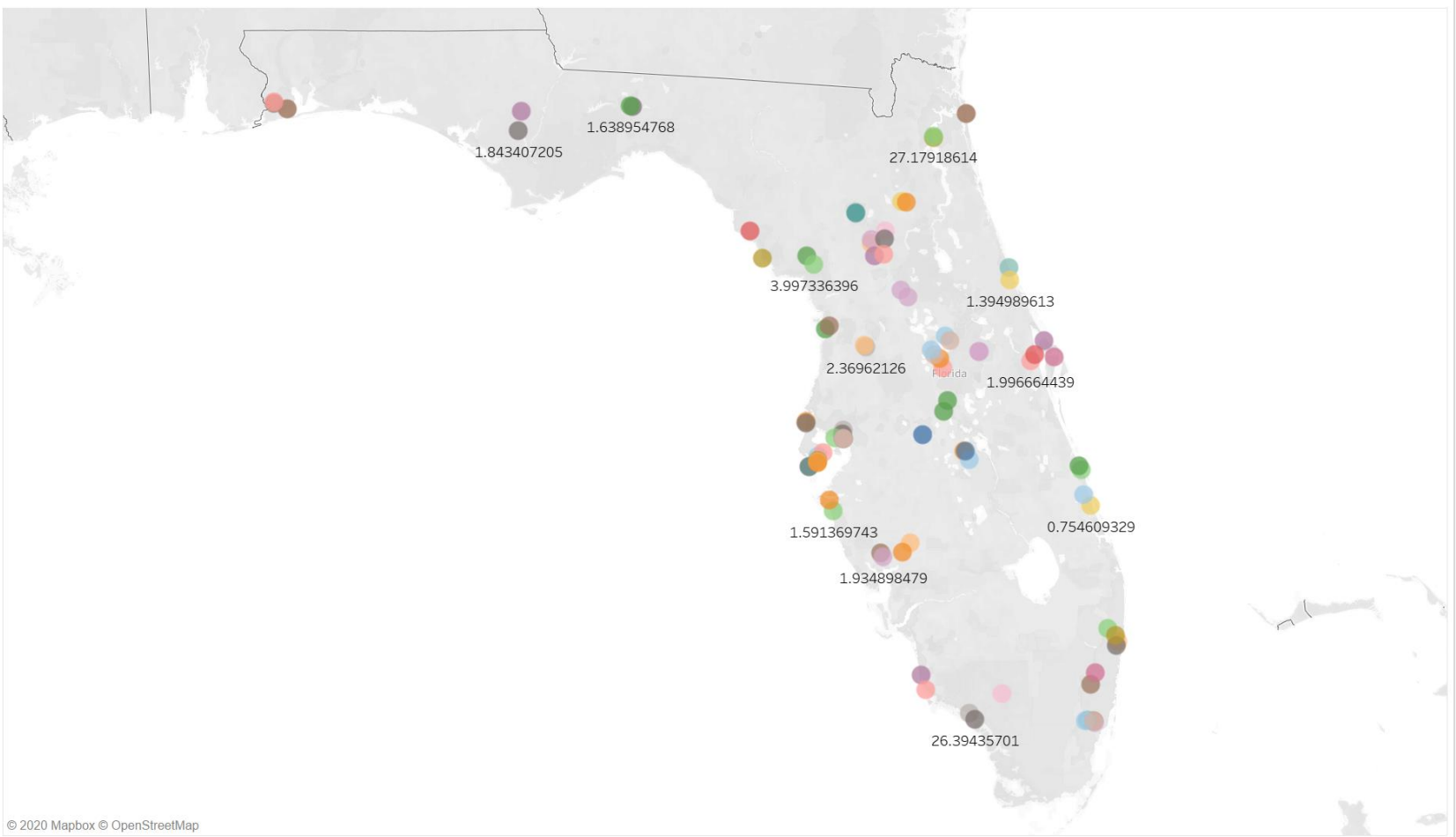# Dimensionality Reduction - Principal Component Analysis

- All the factors were reduced to 5 components
- Component 1 – Wealth & Work
- Component 2 – Type of employment, transport type
- Component 3 – Population size
- Component 4 – Specific ethnicities
- Component 5 – Specific ethnicity

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
ChildPoverty	0.92	0.10	0.10-	0.15-	0.03-
Service	0.89	0.30-	0.11-	0.17-	0.01
Poverty	0.88	0.19	0.06-	0.20-	0.02
Hispanic	0.82	0.11	0.09	0.05	0.20-
Production	0.80	0.41-	0.22-	0.16	0.03-
Unemployment	0.74	0.21-	0.06-	0.42-	0.06
Transit	0.51	0.57	0.06	0.37-	0.11-
Construction	0.48	0.60-	0.23-	0.25	0.13-
MeanCommute	0.47	0.75-	0.12-	0.14-	0.08-
Black	0.32	0.28-	0.07-	0.74-	0.23
Office	0.26	0.53-	0.00	0.20	0.26
Carpool	0.07	0.73-	0.14-	0.49	0.02-
PrivateWork	0.07	0.79	0.06	0.40	0.23
Men	0.00	0.12	0.98	0.05	0.02-
TotalPop	0.01-	0.09	0.99	0.01-	0.01-
Women	0.02-	0.06	0.99	0.05-	0.01-
PublicWork	0.04-	0.87-	0.03-	0.41-	0.06-
Native	0.07-	0.10	0.01-	0.01	0.75
FamilyWork	0.07-	0.08-	0.01	0.20	0.04-
SelfEmployed	0.08-	0.64	0.03-	0.18	0.40-
Asian	0.12-	0.02	0.10-	0.77	0.26
Drive	0.20-	0.90-	0.13-	0.21	0.04-
Citizen	0.24-	0.06	0.95	0.08-	0.01
Walk	0.28-	0.73	0.13	0.08	0.31
Employed	0.29-	0.30	0.89	0.04	0.02
OtherTransp	0.34-	0.72	0.21	0.06-	0.03-
WorkAtHome	0.43-	0.73	0.16	0.14-	0.10-
IncomePerCap	0.73-	0.55	0.20	0.03	0.06
White	0.80-	0.17	0.06	0.23	0.21-
Professional	0.84-	0.48	0.16	0.03-	0.02-
Income	0.85-	0.22	0.11	0.12	0.02

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
ChildPoverty	0.92				
Service	0.89				
Poverty	0.88				
Hispanic	0.82				
Production	0.80				
Unemployment	0.74				
Transit	0.51	0.57			
Construction		0.60-			
MeanCommute		0.75-			
Black				0.74-	
Office		0.53-			
Carpool		0.73-			
PrivateWork		0.79			
Men			0.98		
TotalPop			0.99		
Women			0.99		
PublicWork		0.87-			
Native					0.75
FamilyWork					
SelfEmployed		0.64			
Asian				0.77	
Drive		0.90-			
Citizen			0.95		
Walk		0.73			
Employed			0.89		
OtherTransp		0.72			
WorkAtHome		0.73			
IncomePerCap	0.73-	0.55			
White	0.80-				
Professional	0.84-				
Income	0.85-				

Total market size of BARF is US$ 5.6m in Florida

Total allergic dogs in Florida

1.843407205

1.638954768

27.17918614

3.997336396

1.394989613

2.36962126

Florida

1.996664439

1.591369743

0.754609329

1.934898479

26.39435701

© 2020 Mapbox © OpenStreetMap

Key Takeaway

- On average, **2.78 allergic dogs in Florida** per ZIP code
- **US$ 4,200 revenue potential in Florida per ZIP code**[1]
- **Not a valid business case** to open a dog food shop in a single ZIP code in Florida; however an **online platform delivering to whole Florida** makes a case

---

Further steps required to improve the model

Objective	Worked well	Worked not so well	Next steps
① **Total # dogs in Florida**	• Some predictive power in the demographic data • Principle componenet analysis shows some key drivers of the presence of dogs in a certain neighbourhood	• Not sure if data contains fundamental drivers of dogs • Not sure if NY and FL are comparable	• Analyze the similarities between florida and newyork • Add additional data sources (e.g. dog parks , pet shops etc)
② **Percentage of allergic dogs in Florida**	• Fairly consistent results in predicting above average number of allergic dogs	• Low data volume • Not sure if data contains fundamental drivers of dogs • Not sure if NY and FL are comparable	• Add additional data sources (e.g. dog parks , pet shops etc)