Customer Insights

Classification and Clusterization
Data Analysis

The Company

ONLINE INTERMEDIATE TOURISM AGENCY

FOCUS ON CHEAP/PROMOTION/ERROR FARE PRICES

HIGHLY DEPENDENT ON SOCIAL MEDIA

Process

Online Survey - 1Q 2019

Data collection via Survey Monkey – Market Research Team
3754 responses

Data Analysis – Excel

Reporting - PPT

EXTRA WITH THIS PROJECT:

- Classification model to predict how much users would spend on a trip
- Cluster model to group similar users and understand their behaviors and characteristics
- Visualization in Tableau

Challenges

- Database with almost only categorical variables

- Extensive data cleaning

- Data imbalance

- Visualization

Classification model

- Target: "How much are you willing to spend on a weekend trip for two people?"

- Turned into binary values

- Best model: XGB Classifier

- Scalling: MinMax

- SMOTE

Classification report

	precision	recall	f1-score	support
0	0.76	0.87	0.81	812
1	0.47	0.29	0.36	315
accuracy			0.71	1127
macro avg	0.61	0.58	0.59	1127
weighted avg	0.68	0.71	0.69	1127

The company's focus is on cheap deals and flights and most of its customers look for that in the product. More luxury client is a nice to have, with eventually some deals and packages being directed at those users. I will choose the model that:

1. Has the best recall for 0 (those who are willing to spend up until £199 on a trip), so to be sure that the model is correctly predicting these clients.

2. Has a good recall but also precison for 1, as to have the greatest possible correct prediction from all those predicted 1.

Feature importance

	0	
user_budget_I¬†prefer budget travel, only going for the cheapest deals and offers	0.068090	Budget customers – go for the cheapest deal
accomodation_hotels_5	0.034929	People who prefer to stay at 5 stars hotels
household_income_¬£3200 - ¬£4499	0.030662	People making between 3200 and 4499
accomodation_hotels_1_3	0.024165	People who prefer to stay at 1 to 3 stars hotels
professional_status_Working full-time (fixed location)	0.019188	People working full time
travel_companion_With a partner/spouse	0.017288	People travelling mostly with partner/spouse
travel_motivation_wellness	0.014398	People motivated by wellness when travelling
booking_preference_Always offline (i.e. via travel agency, telephone hotline, etc.)	0.014139	People who prefer to book offline
accomodation_camping	0.013203	People who prefer to camp
accomodation_airbnb	0.012865	People who prefer to stay at airbnb's

Cluster model

PCA – dimensionality reduction

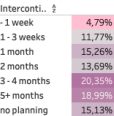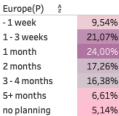MinMax Scaler

Distribution Score Elbow

Kmeans model

Dashboard Tableau
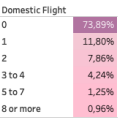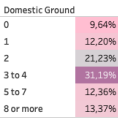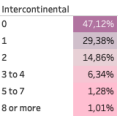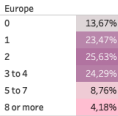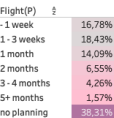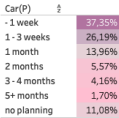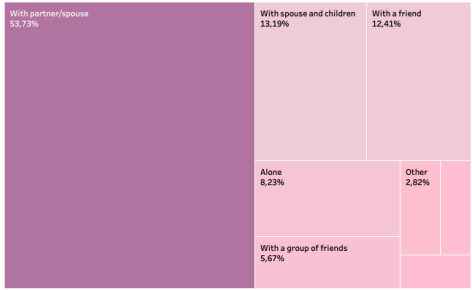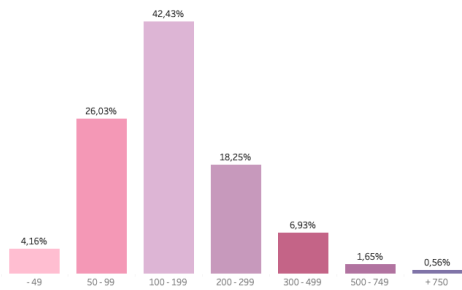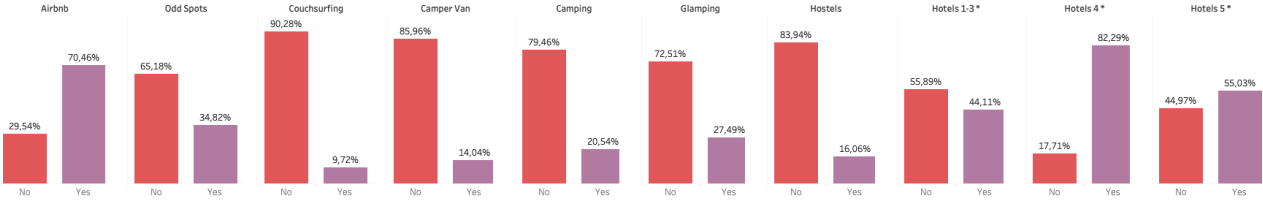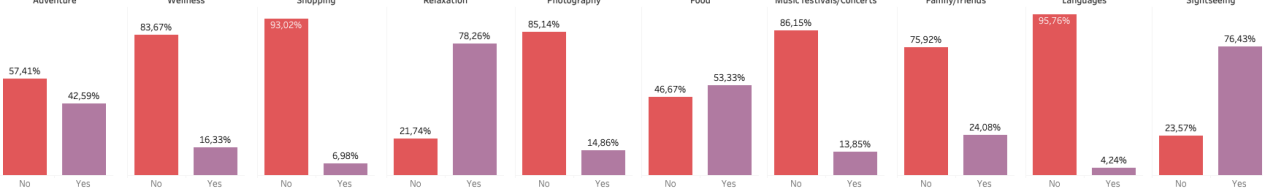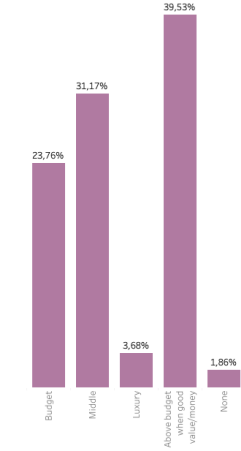
Impact in the Company

01

High: Business and Market Intelligence

02

Medium: Performance, Content and Commercial

03

Low:
Digital Marketing

Thank You