Analytics Start-up Plan

Project	Predicting the spending habits of young individuals in Slovakia
Requestor	Manzano, Wilma Ellaine Rimando, Karon Mae
Date of Request	18-July-2022
Target Quarter for Delivery	
Epic Link(s)	
Business Impact	

Table of Contents

1.0 Business Opportunity Brief

The United Nations reported there are about 1.2 billion young people today which accounts to roughly 16% of the global population (Youth, 2022). Through this data, we can expect a number of opportunities and potential as the young adults, the millennials, tend to have an encouraging mindset towards building a better future. In fact their role today is vital as they are considered to be the main agents of change and progress.

One of the fundamental skills that the young must take into consideration though is having financial literacy as it will benefit them now and in the future in terms of making sound financial decisions. While saving money is an ultimate aspect to financial independence, spending it wisely is of equal importance.

For this project, given the demographics along with their preferences, hobbies, interests and personality traits, we aim to better understand the spending patterns of a young person living in Slovakia, a country considered as one of the youthful countries in Europe. In fact, according to Minns, young people in Slovakia represents 34.17% of their total population (Rural Neets in Slovakia, 2022). Given that this age group represents a high proportion and considered as the next in-line in terms of purchasing power, we aim to provide an insight that will help eCommerce businesses, such as Amazon, as they continue to explore means that will help them engage and win over the hearts and loyalty of the young consumers through marketing and advertising.

1.1 Supporting Insights

The eCommerce industry significantly accelerated and reached its peak at the time of the pandemic as people reverted to online shopping for safety, ease and convenience.

There is no doubt that it will continue to rise over the years for several reasons including: personalization, mobility and convenience, more varieties, and its integration with social media for a wider reach; and so understanding its growth will enable businesses to penetrate into this emerging segment in a more meaningful way.

In 2021, Slovakia was reported to be the 59th largest market for eCommerce with its revenue amounting to $1.4 billion US dollars (ecommerceDB, 2022). Statistics show that there are currently 3.5 million eCommerce users in the country and is expected to grow as new markets and new trends continue to emerge. According to the same report published by ecommerceDB, Fashion seem to be the top segment in Slovakia at 34%, followed by Electronics and Media at 29%, Toys/Hobby/DIY at 18%, Food & Personal Care at 10%, and the remaining percentage accounts for Furniture and Appliances (ecommerceDB, 2022).

Although the presence of the global eCommerce website, Amazon, is available in Slovakia, local eCommerce websites remain to be popular such that the top three stores: alza.sk, mall.sk and itesco.sk contributed to 50% of online revenue in Slovakia in 2021. While the eCommerce business in the country remains to be favourable due to its large and active consumer market, Amazon should explore its full potential by taking into consideration the components of localization where it will lead them to opportunities such as increase in customer engagement and conversion. One way is through understanding the millennials' spending habits better and looking into the best segments they could explore and potentially target.

1.2 Project Gains

Through this study, we aim to better understand several factors that will best explain a young individuals' spending patterns, whether a relation exists or not given their preferences, hobbies, interests and personality traits. The insight that will be derived from this study will be beneficial for Amazon as it will help them optimize their customer service to the fullest, allowing them to enhance their offerings and giving them opportunity to improve their strategies in keeping up with the latest trends through marketing, and more importantly boosting their sales while increasing their online presence and reach in this growing segment through advertising.

Maximizing data to its full potential especially in the eCommerce world is vital because it empowers businesses as they gain relevant insights into customer behaviour and at the same time opens up a number of opportunities for growth. It would be ideal for businesses to utilize data to its highest capacity so as to prevent any missed opportunities leading towards progression.

2.0 Analytics Objective

We would like to test whether demographics along with certain preferences, hobbies, interests, and personality traits are good predictors of a young adult's spending habits. Specifically, the null and alternative hypothesis are as follows:

Null Hypothesis (*Ho*): There is no significant relationship between a young adult's spending habits and his/ her music preferences, movie preferences, hobbies, interests, and personality traits.

Alternative Hypothesis (*Ha*): There is a statistically significant relationship between a young adult's spending habits and his/ her music preferences, movie preferences, hobbies, interests, and personality traits.

During the analysis, we also would like to answer if certain number of predictors can be summarized into smaller representative variables to simplify our model and make it easier to interpret and visualize. Lastly, to aid stakeholders in meeting their objective in understanding the young adult population better, we will delve into segmenting and characterizing them to find out the best segments that Amazon could potentially target.

2.1 Other related questions and assumptions:

We would also seek understanding on the following question/s:

1. Are there significant differences in spending habits and other preferences between age groups (15 to 20 years old versus 21 to 30 years old)?

We take note of some assumptions that we need to consider in our analysis and interpretation.

1. Survey data was not sampled randomly, rather through convenience and snowball only. Due to this, the sample distribution may not reflect actual population census.
2. Spending patterns are based on Likert scale (1 as strongly disagree, and 5 as strongly agree), instead of actual purchase spending.

2.2 Success measures/metrics

We have established several metrics to determine the performance of our analysis such as:

1. Overall goodness of fit of regression model: *high (at least 20%) r-square, low average squared error, accuracy, f-score and their gap between training and validation dataset*
2. Number of independent variables: *least number of variables that has the most optimal fit is desired*
3. Number of clusters identified: *4 to 5 customer segments that can be easily targeted*

2.3 Methodology and Approach

Type of Analysis:

The initial approach will be to utilize data reduction technique, either ***Principal Component Analysis or Exploratory Factor Analysis,*** so we can further reduce the number of predictors and simplify our predictive model. Next, we will use ***regression*** to determine variables that are significant predictors of spending habits of young individuals. Our third step would be employing ***cluster analysis*** to determine customer segmentation and their specific characteristics.

Methodology:

We will start the analysis by data gathering and exploration and identifying variables that would best correspond to our primary target, young individual's spending as well as potential predictors. During the exploratory stage, we will also investigate variable distribution and find data patterns that would affect our analysis. We will also

implement required data cleaning and wrangling, such as recoding of out-of-range values and converting each variable to their correct data types to become feasible for modelling. The final dataset for analysis will be split equally between training and testing dataset, where the training dataset will be used for the principal component analysis and regression modelling and the results will be further validated on the test data. We will then explore segmentation through cluster analysis and run cross tabulations of the determined clusters against other variables to understand their behaviour and characteristics.

Output:

Our primary output would be the set of factors or independent variables that would best explain spending patterns of young individuals. The cluster groups are also essential for costumer profiling. Based on these results, we will provide actionable insights and recommendations on what Amazon could do to reach out more to young consumers in the form of advertising, marketing messages, and promotional offers. We will tackle strategies on how Amazon could utilize these results to increase their penetration and market share.

3.0 Population, Variable Selection, Considerations

Audience/population selection:	Students of the Statistics class at FSEV UK were asked to invite their friends to participate in a survey. The aim of the survey was to explore the different preferences, interests, habits, opinions, and including fears of young adults in Slovakia.
Observation window:	The survey was conducted in 2013 and was presented in both electronic and written form.
Inclusions:	The participants of the survey were of Slovakian nationality and aged between 15-30 years old.
Exclusions:	None Reported
Data Sources:	Our study was based on data derived from Kaggle.com on "Young People" in Slovakia. https://www.kaggle.com/datasets/miroslavsabo/young-people-survey The original questionnaire was in Slovak language and was later on translated into English.
Audience Level:	The insight that will be derived from this study will be beneficial for eCommerce businesses, such as Amazon, as it will help them optimize their customer service to its full potential through gaining relevant insights into customer behaviour and allowing them to work on a number of opportunities for growth.
Variable Selection:	For this study, the young individuals demographics along with certain preferences, hobbies, interests, and personality traits will be tested whether they are good predictors of a young adult's spending habits. Our goal is to find independent variables that would best explain spending patterns of young individuals.
Derived Variables:	We derive spending patterns by taking mean aggregate of their individual spend rating on several items – shopping centres, branded clothing, socializing, appearance, gadgets, and food. In addition, the segmentation resulting from cluster analysis will be cross tabulated with other variables to understand their behaviour and characteristics for costumer profiling.
Assumptions and limitations:	According to Act No. 282/2008 on youth work support, a youth in Slovakia is a person who is up to 30 years old. For this project, we focused on the 15-30 age range.

4.0 Dependencies and Risks

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
Since the survey was conducted through snowball sampling where students were asked to invite their friends to participate in the survey, the outcome of the project may be influenced because the result may not reflect the actual population census.	*Low*	-	*The scale used in the questionnaire were rated from 1 as Strongly Disagree to 5 as Strongly Agree.* *As part of our initial analysis, we generated frequency tables to check whether the responses were concentrated in one area or are spread out.* *Moreover, we also checked the skewness and kurtosis of our data. The result after our initial check were all within the acceptable range.*

5.0 Deliverable Timelines

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Project Kick-off	Phase 1: Project Planning	▪ Business Objective ▪ Final Dataset ▪ Research Questions ▪ Establishing hypothesis ▪ Tools and Platforms to Use	1 day	July 8, 2022
2.	Data Exploration and Analysis	Phase 2: Data Exploration	▪ Variable identification ▪ Variable distribution (skewness) in SPSS ▪ Data cleaning (out-of-range values, data types) in SPSS ▪ Correlation Check	5-7 days	July 11-15, 2022
3.	Modelling	Phase 3: Modelling	▪ Python Coding ▪ Principal Component Analysis ▪ Regression ▪ Cluster Analysis ▪ Results Interpretation	10-14 days	July 17-29, 2022
4.	Reporting	Phase 4: Report Production	▪ Written report ▪ Governance ▪ Documentation	3-5 days	July 30-Aug 3, 2022
5.	Story Board	Phase 4: Report Production	▪ Presentation Outline	1-2 days	August 4-5, 2022
6.	PowerPoint material production	Phase 4: Report Production	▪ Initial Presentation Draft	3-4 days	August 6-9, 2022
7.	Peer Feedback	Phase 4: Report Production	▪ Mock presentation ▪ Initial feedback (peer) and	1 day	August 10, 2022

			suggested revisions		
8.	PowerPoint revisions	Phase 4: Report Production	▪ Revised PowerPoint material	2-3 days	August 11-12, 2022
9.	Final Presentation	Phase 5: Report Presentation	▪ Final presentation to stakeholders ▪ Stakeholders' comments and feedback	1 day	August 15, 2022
10.	Written report/ PowerPoint revisions 2	Phase 4: Report Production	▪ Revised PowerPoint material	2-3 days	August 16-18, 2022
11.	Delivery & sign-off	Phase 6: Approval	▪ Final approval	1 day	August 19, 2022

References

United Nations. (17 July, 2022). *Youth*. Retrieved July, 2022, from United Nations: https://www.un.org/en/global-issues/youth

Minns, A. (18 July, 2022). *Rural Neets in Slovakia*. Retrieved from rnyobservatory.eu: https://rnyobservatory.eu/web/National-Reports/NR-SLOVAKIA-09-19.pdf

ecommerceDB. (18 July, 2022). *The eCommerce Market in Slovakia*. Retrieved from ecommerceDB: https://ecommercedb.com/en/markets/sk/all