School of Analytics
Admission Case Study 2023/24

Assignment Overview

Dear applicants,

Congratulations on successfully passing the admission test!

The next step in the admission process is solving a case study. This assignment should provide you with a brief insight into typical tasks and responsibilities that a Data Analyst faces every day when making data-driven decisions and recommendations to solve real-life business problems.

Please see the case outline and further instructions below.

Good luck!

Sincerely Yours,

School of Analytics Team

General Remarks

The case study is designed to test your ability to identify the key features and attributes of a given set of data, determine the underlying relationships and apply the data analysis process to derive meaningful conclusions.

Successful applicants are expected to demonstrate strong analytical and problem-solving skills and a good command of fundamental data analytics tools.

Software Requirements

Although there are no specific software requirements to complete the assignment, we strongly recommend using the following applications:

- MS Excel (preliminary data analysis)
- Python/R/ Stata (advanced[1] data analysis)
- MS PowerPoint (results presentation)

[1] Please note that using programming languages is optional since basic data analysis can be performed in MS Excel. However, it would be an advantage to showcase your technical skills by performing a more complex data analysis with Python, R or Stata.

Case Study Description

For the main assignment, please read the description, problem outline and additional information for the “Marketing Analytics” case study.

The case study contains the following sections:

- Introduction
- Problem Outline
- Data and Additional Information
- Task and Questions Outline
- Metadata

Marketing Analytics

Introduction

You work as a Data Analyst for a mid-range and premium home improvement company. The company's primary sales region covers Moscow and Moscow Oblast, and recently a new delivery option was added which allows to reach customers across the whole country.

Problem Outline

According to the recent customer review analysis, there was a drop in the share of middle-aged customers (35–59 years old) who tend to be the most profitable customers for the company. By analysing the preference pattern of customers within this age category, your task as a Data Analyst is to provide recommendations for a successful advertising campaign targeting this specific demographic.

Data and Additional Information

The company has conducted 4 customer surveys to gather information about the customers' overall online shopping experience and preferences. The survey details are outlined below.

Survey	Description
Goods review	• gender • age • first purchase ("Yes" or "No") • service rating • product rating
Delivery service review	• gender • age • service rating • product rating
Favourite products	• gender • age • bedding, bathroom products, blankets, kitchen products, fragrances and decor, home clothes ("Yes" – have bought the product before, "No" – have never bought the product)
Category	• gender • age • bedroom, bathroom, kitchen, dishes, home clothes, curtains (1 – "Unlikely to purchase the product", 5 – "Highly likely to purchase the product")

Task and Questions Outline

1. Define a business problem
2. Formulate a goal and describe how you will achieve it
3. Describe the data that will be used in the analysis. What additional sources of information can be used?
4. Define data analysis methods
5. Analyse the data:
    a. Describe the data. What conclusions can be drawn in regards to the planning of an advertising campaign?

   b. Determine the needs of the target audience. What do you need to pay attention to?
   c. Visualise the data using various methods (line chart, column chart, etc.). Please use at least 3 different chart types
6. Create a PowerPoint presentation to summarize and showcase the key findings

Metadata

The file **"market_poll1.csv"** contains data related to product review:

#	Name	Description
1	sex	*gender of the interviewee*
2	age	*age of the interviewee*
3	first_purchase	*1 if this is the first purchase, otherwise 0*
4	grade_service	*service rating from 1 (very bad) to 5 (very good)*
5	grade_product	*product rating from 1 (very bad) to 5 (very good)*

The file **"market_poll2.csv"** contains data related to delivery service review:

#	Name	Description
1	sex	*gender of the interviewee*
2	age	*age of the interviewee*
3	first_purchase	*1 if this is the first purchase, otherwise 0*
4	grade_service	*service rating from 1 (very bad) to 5 (very good)*
5	grade_product	*product rating from 1 (very bad) to 5 (very good)*

The file **"market_poll3.csv"** contains data related to customer product preferences:

#	Name	Description
1	sex	*gender of the interviewee*
2	age	*age of the interviewee*
3	bed_linen	*1 if the interviewee bought bed linen, otherwise 0*
4	bath	*1 if the interviewee bought bathroom products, otherwise 0*
5	blanket	*1 if the interviewee bought blankets, otherwise 0*
6	kit	*1 if the interviewee bought goods for the kitchen, otherwise 0*
7	decor	*1 if the interviewee bought fragrances and decor, otherwise 0*
8	cloth	*1 if the interviewee bought clothes for the home, otherwise 0*

The file **"market_poll4.csv"** contains data related to product category preferences:

#	Name	Description
1	sex	*gender of the interviewee*
2	age	*age of the interviewee*
3	bed	*the likelihood of buying bedroom goods from 1 (low likelihood) to 5 (high likelihood)*

4	bath	*the likelihood of buying bathroom goods from 1 (low likelihood) to 5 (high likelihood)*
5	kit	*the likelihood of buying kitchen goods from 1 (low likelihood) to 5 (high likelihood)*
6	dish	*the likelihood of buying dishes from 1 (low likelihood) to 5 (high likelihood)*
7	cloth	*the likelihood of buying home clothes from 1 (low probability) to 5 (high probability)*
8	curtains	*the likelihood of buying curtains from 1 (low probability) to 5 (high probability)*

Extra Credit Assignment

Introduction

To earn additional points toward the total grade for the Case Study, applicants are offered to complete the Extra Credit Assignment.

This would require the knowledge and application of more advanced data analysis techniques including Natural Language Processing (NLP).

Problem Outline

The main goal of your research is to determine whether the retailer description can be used as an inference for a price category. In other words, you need to explore the relationship between how a particular retailer is described and what price segment they represent.

Data and Additional Information

The dataset contains information about approximately 2740 retailers that operate in Russia. Data was scrapped from a Russian website which collects information about Russian retail real estate and shopping malls.

Task and Questions Outline

1. Import all Python modules you may need: pandas as pd, numpy as np, nltk, re, sklearn
2. Import the data from the "russian_retail.csv" file: pd.read_csv()
3. To convert texts into feature sets using the bag-of-words method, you must first filter the text, leaving only meaningful words. Download a set of insignificant words from the corpus for filtering purposes with the nltk.download() command (in the window that appears, Corpora tab, stopwords item)
4. Import stopwords from nltk
5. Define filter functions
6. Copy data using copy() function into a new variable and use filter function .apply(lambda s: smth_to_words(s)). Print the first three rows
7. Visualise regions using WordCloud
8. Visualise description using WordCloud
9. Visualise other columns with any charts. Why did you use this type of visualisation?

For a more detailed task description and a step-by-step guideline, please refer to the Jupyter Notebook template **"SoA – Admission Case Study – Extra Credit Assignment".**

Metadata

The file **"russian_retail.csv"** contains 7 columns that identify each retailer by specialization, price category, regions of presence and a comprehensive description of its business.

Name	Data Type	Description
name	object	*# retailer name*
country_origin	object	*# retailer's country of origin*
domain	object	*# business specialization*
price_category	object	*# price category*
founded	int64	*# the year the business was founded*
presence_regions	object	*# cities and Russian regions where the retailer has outlets*
description	object	*# retailer description*

Submission Guidelines

Below you can find the general submission requirements for the case study assignment.

Your submission should include the following files:

Main Assignment

1. MS PowerPoint presentation converted to a PDF file
   - File name: “Surname Name – Presentation – Marketing Analytics – 2”
   - A presentation should have **NO MORE** than 2–3 slides[2]
2. MS Excel spreadsheet **OR** Python/R/Stata file
   - File name: “Surname Name – Data Analysis – Marketing Analytics – 2”

Extra Credit Assignment

1. Jupyter Notebook with Python code
   - File name: “Surname Name – Extra Credit Assignment – 2”

All files must submitted no later than **23:55 September 29**.

Submissions received after the deadline will not be accepted or marked.

[2] Please note that any information outside the maximum slide limit will not be reviewed or counted toward the final grade.

Grading Rubric

Grading criteria for the Main Assignment are summarised in Table 1.

Table 1. Case Study Grading Rubric			
#	Criteria	Description	Points
		MAIN ASSIGNMENT	
1	Business Problem	The business problem is clearly stated and matches the problem outline	2
2	Research Design	The goal and research process are clearly outlined and correspond to the problem outline	2
3	Data Description	The key features of the dataset are correctly identified and described	2
4	Methodology	A detailed description with a list of data analysis methods is provided	3
5	Data Analysis	A preliminary data analysis is performed: • Data importing • Data aggregation and transformation • Data visualisation • Data interpretation	4
6	Results	Results are clearly articulated and correctly inferred from the data analysis performed	3
7	Recommendations	Recommendations are clearly outlined and correspond to the results	2
8	Style and Formatting	A submitted file has a clear structure and adheres to the general standards for document formatting: • Data analysis in Excel has a unified format across the entire spreadsheet (numbers, fonts, alignments, etc.) **OR** • Data analysis completed in Python, R or Stata follows style guides for the specific programming language	1
9	Presentation	A submitted file has a clear structure and adheres to the general standards for document formatting: • Slides formatting in MS PowerPoint (numbers, fonts, alignments, etc.) • MS PowerPoint slides are concise and not overloaded with information • MS PowerPoint presentation has a clear and logical information flow	1
	Total		**20**

Grading criteria for the Extra Credit Assignment are summarised in Table 2.

Table 2. Case Study Grading Rubric

#	Criteria	Description	Points
		EXTRA CREDIT ASSIGNMENT	
1	Exploratory Data Analysis (EDA)	A preliminary data analysis is performed: • Data importing • Data cleaning • Data aggregation and transformation • Data visualisation • Data interpretation	2
2	Natural Language Processing (NLP)	Relevant Python libraries for natural language processing are used to perform text analysis and draw conclusions from the data	4
3	Visualisation	All required visualisations are performed using the Python wordcloud library and other visualisation types	3
4	Style and Formatting	A submitted file has a clear structure and adheres to the general standards for document formatting: • Data analysis is completed in Python and the solution is submitted as a Jupyter Notebook • The code is thoroughly commented on and follows PIP 8 style guidelines	1
	Total		**10**
	Total inc. extra credit		**30**