**Assignment 1: Prestige Mall**

|  |  |
| --- | --- |
| Name | **WONG QI YUAN JEFFREY** |
| ID | **P7359567** |
| Class | **PA-01** |

**Case Study: Prestige Mall**

**What businesses have you here?**



One of the shops in this shopping centre, *Prestige Mall*, has become vacant. You were tasked by your boss to suggest a possible tenant to bring into the Mall.

Some decisions that we make may be based on personal judgments but some may not. In this case, a proposal with support from data collected is obviously more convincing as compared to a proposal without such a support or a proposal supported with personal judgments alone. What are some of the information that you would include in your proposal, as support to the kind of business (and thus the potential tenant) that you would like to recommend to your boss?

Well, for instance, you may like to have an initial “feel” of the profile of customers who patronize this mall:

* What kind of job sectors are they from?
* How many times do they frequent Prestige Mall in a month?
* How much do they typically spend in Prestige Mall?
* What is their age profile?
* What is the proportion of male customers of Prestige Mall?
* What is their average monthly household income?

The questions listed above would generate valuable information in helping you to decide on the potential tenant that you would like to recommend to your boss.

You will like a systematic way to approach this problem, thus you use the 4-steps **statistical problem-solving process** that you have been introduced:

Formulating  
Questions

Collecting  
Data

Analysing  
Data

Interpreting  
Results

In order to collect data, a selected group of customers of Prestige Mall is chosen to respond to a survey. By the end of the survey period, you have collected feedback from 200 customers. The data is recorded in the spreadsheet named *PrestigeMall\_Data.xlsx*.

You can refer to Q5 below on how to fill in the fields in the table that is designed in accordance to the 4-steps statistical problem-solving process.

Refer to *PrestigeMall\_Data.xlsx* and use Minitab to analyze the data.

Q1: What is the aim of this case study?

|  |
| --- |
| To recommend a potential tenant or business into the Prestige Mall |

Q2: What is the sample of this case study? And what is the targeted population?

|  |
| --- |
| Sample: Prestige Mall customers who took the survey  Population: All customers who visit Prestige Mall. |

Q3: How were the data collected, as recorded in the data file named “Prestige Mall”?

|  |
| --- |
| Data were collected through customers responses (i.e. surveys) for the selected variables and recorded into Spreadsheets for Data Analysis using Statistical Software (i.e. Minitab 18) |

Q4: What information (variables) does the data file named “Prestige Mall” hold?

|  |
| --- |
| Job Sector, Age, Number of Visits Per Month, Gender, Household Income, Amount Spent Per Month |

Formulating Questions

Collecting Data

Analyzing Data

Interpreting Results

|  |  | **Define variable and type of data** | **Both Numerical and Graphical summaries to be included** | **Description of Sample and  Generalization to the target population**  **(Based on Q5-Q7 sample description)** |
| --- | --- | --- | --- | --- |
| Q5 | What is the proportion of male and female customers of Prestige Mall? | *Which variable would you use to answer this question?*  Gender  *What type of data are these?*  Qualitative (nominal) | **Male:** 43%  **Female:** 57% | ***Describing the sample:***  **For Q5:** There is a slightly higher proportion of female customers visiting Prestige Mall compared to male customers.  **For Q6:**  Based on the numerical and graphical summaries, the distribution is unimodal and slightly positively skewed to the right (i.e. Mean > Median) with a mean age of 41.955 years old and a standard deviation of 11.861 years old. There is a peak with the highest frequency of aged around 40 to 45 years old. From the boxplot diagram, there are no outliers found and the median age is calculated to be 41.5 years old, this means that half of the customers are aged above 41.5 years old and the others half are aged below 41.5 years. The spread is wider (measured by the box). Hence, the possible customers visiting Prestige Mall are likely to be around 30 to 45 years old.  Additionally, from the histogram, an extreme small number of customers with age around 15 to 20 years old visiting the Prestige Mall, and a possible reason could be due to still schooling.    **For Q7:**  Based on the numerical summary and a histogram, the distribution is unimodal and is slightly positively skewed to the right (i.e. Mean > Median) with a mean household income of $8231.3 and a standard deviation of $1222.9. There is a peak showing the highest frequency of household income of the customers with about $8000 to $8500. Additionally, there is an extreme observe value found on the histogram with household income about $12000 to $12500, possibly an outlier.  Moreover, from the boxplot diagram, there are two outliers (mark with asterisks) found (with household income of $12000). The median is a better measure of the typical household income by these 200 customers. This is because the mean is an arithmetic average and gets affected by the two extreme observations. The standard deviation also gets affected by the two outliers. Both the median and IQR do not get affected by outliers as much.  The spread is narrow (measured by the box) and the customers are likely to have household income with about $7500 to $8500.  ***Generalizing to the population:***  Generally, customers of Prestige Mall are slightly more likely to be female, with aged around 30 to 45 years old, and with household incomes of about $7500 to $8500 per month. |
| Q6 | What is the age profile of the customers? | *Which variable would you use to answer this question?*  Age  *What type of data are these?*  Quantitative (Discrete) | **Numerical Summaries:**    **Graphical Summaries:**        **Mean: 41.955**  **Median: 41.500**  **SD: 11.861** |
| Q7 | What is the distribution of income of the customers? | *Which variable would you use to answer this question?*  Household Income  *What type of data are these?*  Quantitative (Continuous) | **Numerical Summaries:**    **Graphical Summaries:**        **Mean: 8231.29**  **Median: 8067.36**  **SD: 1222.9** |

Formulating Questions

Collecting Data

Analyzing Data

Interpreting Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Define variable and type of data** | **Both Numerical and Graphical summaries to be included** | **Description of Sample and  Generalization to the target population**  **(Based on Q8-Q10 sample description)** |
| Q8 | What is the proportion of customers in the IT/Eng and Bus/Fin sectors? | *Which variable would you use to answer this question?*  Job Sector  *What type of data are these?*  Qualitative (Nominal) | **Graphical Summaries:**    **Bus/Fin: 29.5%**  **IT/Eng: 51.5%**  **Others: 19.0%** | ***Describing the sample:***  **For Q8:**  From the Pie Chart, about half of the customers visiting the Prestige Mall are likely to be from IT/ Eng sectors as compared to both Bus/Fin and others sectors.  **For Q9:**  Based on the bar chart, there is no much difference and the frequency with approximately about 30 to 40 customers across the number of visits per month Even though “1 time of visit per month” has the highest frequency over the others, but it does not affect the overall data.  From the numerical summary, the distribution is slightly negatively skewed to the left (i.e. Mean < Median) with a mean number of visits per month of 3.465 and a standard deviation of 1.742.  Moreover, from the boxplot diagram, there are no outliers found and the median is 4 times of visits per month. This means that there is about 50% of the customers visiting the Prestige Mall with more than 4 visits per month, and the others visit less than that.  Additionally, the spread is wider (measured by the box) and the customers are likely to visit Prestige Mall with about 2 to 5 times of visits per month.  **For Q10:**  Based on the numerical summary and the histogram, the distribution is unimodal and is slightly negatively skewed to the right (i.e. Mean < Median) with a mean amount spent per month of $501.20 and a standard deviation of $137.84. There is a peak showing the highest frequency of amount spent per month by the customers with about $500 to $550. Additionally, there is no obvious extreme data found on the histogram.  Moreover, from the boxplot diagram, there are four outliers (mark by asterisks), that is, well below the minimum amount spent per month of $200. The median is a better measure of the typical household income by these 200 customers. This is because the mean is an arithmetic average and gets affected by the two extreme observations. The standard deviation also gets affected by the four outliers. But, both median and IQR do not affect by outliers as much.  The spread is narrow (measured by the box) and the customers are likely to spent about $400 to $550 per month.  ***Generalizing to the population:***  Generally, customers of Prestige Mall are likely to be from IT/Eng sector, with about 2 to 5 visits per month and spent about $400 to $550 |
| Q9 | How often do the customers visit Prestige Mall in the last month? | *Which variable would you use to answer this question?*  Number of Visits Per Month  *What type of data are these?*  Quantitative (Discrete) | **Numerical Summaries:**    **Graphical Summaries:**        **Mean: 3.465**  **Median: 4.000**  **SD: 1.742** |
| Q10 | How much did the customers spent last month at Prestige Mall? | *Which variable would you use to answer this question?*  Amount Spent Per Month  *What type of data are these?*  Quantitative (Continuous) | **Numerical Summaries:**    **Graphical Summaries:**        **Mean: 501.20**  **Median: 503.11**  **SD: 137.84** |

Formulating Questions

Collecting Data

Analyzing Data

Interpreting Results

|  |  | **Define variable and type of data** | **Both Numerical and Graphical summaries to be included** | **Description of Sample and  Generalization to the target population**  **(Based on Q11 sample description)** |
| --- | --- | --- | --- | --- |
| Q11 | Is there any preliminary evidence to claim that female customers who went to Prestige Mall  6 times last month spent very much more than female customers who went to Prestige Mall only once last month? | *Which variable would you use to answer this question?*  Amount Spent Per Month  *What type of data are these?*  Quantitative (Continuous) | |  |  |  | | --- | --- | --- | | Amount spent/ month | Visit the mall once / month | Visit the mall  6 times / month | | *n* | 24 | 19 | | Mean | 470.4 | 492.2 | | SD | 155.8 | 127.2 | | Minimum | 130.0 | 271.5 | | Q1 | 398.6 | 406.3 | | Median (Q2) | 454.4 | 489.3 | | Q3 | 563.2 | 572.1 | | Maximum | 743.8 | 749.5 |   **Numerical Summaries:**    **Graphical Summaries:** | ***Describing the sample:***  From the boxplots diagram and numerical summaries, the mean, median and the spread (as measured by the “box”) of amount spent per month for female customers visiting the mall once per months and for female customers visiting the mall for 6 times per month are approximately the same. The minimum amount spent for female customers visiting the Prestige Mall for six times per month is generally higher than that of female customers visiting the Prestige Mall for once per month, but the maximum amount spent per month for female customers visiting the Prestige Mall is also approximately the same for both cases.  However, there is an outlier of amount spent per month found for female customers visiting the mall once per month, but no outlier found for female customers visiting the mall six times per month. In fact, the outlier has greatly affected both the mean and standard deviation for female customers visiting the mall once per month. But, both the median and IQR are not greatly affected by the outlier. Thus, both are good measure for the variation and for both cases the variation is about the same.  ***Generalizing to the population:***  There is no evidence to claim that female customers who went to Prestige Mall 6 times last month spent very much more than female who went to Prestige Mall only once last month. |

Formulating Questions

Collecting Data

Analyzing Data

Interpreting Results

|  |  | **Define variable and type of data** | **Both Numerical and Graphical summaries to be included** | **Description of Sample and  Generalization to the target population**  **(Based on Q12 sample description)** |
| --- | --- | --- | --- | --- |
| Q12 | **Ask a question about the relationship between 2 variables, then proceed to investigate.**  **Is there a relationship between Customers Age and Household Income?** | *Which variables would you use to answer this question?*  Customers Age (Explanatory)  Household Income (Response)  *What type of data are these?*  Quantitative (Continuous) | **Numerical Summaries:**    **Graphical Summaries:** | ***Describing the sample:***  From the scatter plot, there is no correlation and association between the customers age and the household income since the Pearson correlation is approximately close to 0.    ***Generalizing to the population:***  Generally, customers from Prestige Mall comes from different age groups have no influence on their household incomes. |

Q13: Based on interpretations from Q5 to Q12, suggest a potential business to recommend to your boss, and thus a potential tenant (e.g Rolex?).   
 Note: You need not use all the interpretations.

|  |
| --- |
| Based on all interpretations from Q5 to Q12, I would suggest that Fashion (i.e. clothing, shoes and accessories) would be the potential business in the Prestige Mall, which are generally suitable for professional working adults, and the potential tenant would be H&M. |

Q14: State three assumptions that may be applicable in this case study.

|  |
| --- |
| * Assumed that all respondents involved in this case study gave accurate and honest responses, otherwise the study would become irrelevant. * Appropriate selection of sampling technique (i.e. random sampling, etc) and sample must be independent (i.e. the value of one observation does not influence the value of other observations) * Appropriate selection of graphical and numerical summaries based on the formulating questions. |

Q15: Given a chance to redesign the survey, suggest a possible new question (and thus a new variable) that can be added into the survey to gain more insight on the customers of Prestige Mall. Explain how this variable may affect your recommendation in Q13.

|  |
| --- |
| Formulating Question:  What is the highest academic qualification do the prestige customers completed?   * PSLE * N/ O Levels * Diploma * Bachelor’s Degree * Postgraduate   Collecting Data:  *What is the name of the variable you would use to answer this question?*  Academic Qualification  *What type of data is this?*  Qualitative (Ordinal)  Analyzing Data:  *What numerical summaries do you need?*  Not Applicable  *What type of graphical summary do you need (i.e. pie chart, bar chart, histogram, boxplot?). Explain why you use this type of graphical summary.*  Bar Chart – It generally gives a better visual overview of the frequency between categories  Explain how this variable may affect your recommendation in Q13.  Respondents who are well educated, especially with Bachelor’s degrees and above, are therefore to have acumen and  pedigree to make prudent choices and response on the survey, so academic qualification is very crucial to be included in the  survey questionnaire. The statistical analysis would therefore be made more appropriate, which in turn to provides more  relevant recommendation of the potential business to the Prestige shopping mall. |