Chapter 5
Demographics and
Delhi Metro Rail Corporation
CHAPTER 5
DEMOGRAPHICS AND DELHI METRO
RAIL CORPORATION

5.1. AN OVERVIEW
5.2. DESCRIPTIVE STATISTICS
5.3. DESCRIPTIVE STATISTICS FOR DEMOGRAPHY (AGE AND GENDER)
   5.3.1 Gender - Graphical Presentation
   5.3.2 Age - Graphical Representation
5.4 TEST OF NORMALITY
5.5 SUMMARY OF CHI-SQUARE – AGE AND GENDER
   5.5.1 Gender
   5.5.2 Age
The Research was carried out in National Capital region of India (Delhi, Ghaziabad, Faridabad, Noida and Gurgaon). The research gave a first hand experience of understanding the mentality of commuters. Demographics refer to the statistical data of a population, especially those showing average age, income, education. For the present research the two demographic factors Namely, Age and Gender of the commuters have been focused by studying the relationship between them and overall perception of Delhi Metro Usage. Test of normality was applied to ascertain that data was normally distributed. The Chi Square test was applied to ascertain the association between the demographic variables and overall perception of metro usage. The test revealed that there was a significant relationship between Gender and overall perception of usage of Delhi Metro. The test also ascertained that there was significant relationship between the demographic variable, Age and overall perception of Delhi Metro Usage.
CHAPTER 5
DEMOGRAPHICS AND DELHI METRO RAIL CORPORATION

5.1. AN OVERVIEW

The Research\(^1\) was carried out in National Capital region of India (Delhi, Ghaziabad, Faridabad, Noida and Gurgaon). Sincere attempt has been made by the researcher to cover as area wide as possible, so as to achieve the objective of the study.

A sample of 1200 respondents was taken into consideration. These respondents were interviewed through a well drafted pre-tested questionnaire. The respondents were very forthcoming and cooperated for most part. That was reason why, the response rate was 90%, out of 1200, 1080 completely filled questionnaires.

The survey conducted provided loads of experiences. However, there were some difficulties encountered while performing the survey. One of the main problems faced while conducting the survey was language. Communication posed to be a problem in some cases as the questionnaire was in English.

In some cases contradictory answerers were also given which seemed to show that the respondents were trying to get over the interview as soon as possible.

Not much of importance was attached to the survey by some of the respondents. The researcher was either greeted with reluctance or nonchalance. In some cases

contradictory answerers were also given which seemed to show that the respondents were trying to get over the interview as soon as possible. Whilst the benefits of secondary sources are considerable, their shortcomings have to be acknowledged. There is a need to evaluate the quality of both the source of the data and the data itself. The main problems may be categorized as follows:

The researcher had to be careful, when making use of secondary data, of the definitions used by those responsible for its preparation. It should be noted that definitions may change over time and where this is not recognized erroneous conclusions may be drawn.

When a researcher conducts fieldwork she/he is possibly able to estimate inaccuracies in measurement through the standard deviation and standard error, but these are sometimes not published in secondary sources. The only solution is to try to speak to the individuals involved in the collection of the data\(^2\) to obtain some guidance on the level of accuracy of the data. The problem is sometimes not so much ‘error’ but differences in levels of accuracy required by decision makers.

The Researcher has to be very vary of the vested interests when he consults secondary sources. Those responsible for their compilation may have reasons for wishing to present a more optimistic or pessimistic set of results for their organization. The reliability of published statistics may vary over time. It is not uncommon, for instance, for the systems of collecting data to have changed over time but without any indication of this to the reader of published statistics.

---

Geographical or administrative boundaries may be changed by government, or the basis for stratifying a sample may have altered. Other aspects of research methodology that affect the reliability of secondary data are the sample size, response rate, questionnaire design and modes of analysis. Most censuses\(^3\) take place at 10 year intervals, so data from this and other published sources may be out-of-date at the time the researcher wants to make use of the statistics. The time period during which secondary data was first compiled may have a substantial effect upon the nature of the data. Whenever possible, marketing researchers ought to use multiple sources of secondary data. In this way, these different sources can be cross-checked as confirmation of one another. Where differences occur an explanation for these must be found or the data should be set aside.

A special care was taken to maintain the equality in respondent’s demography. Not much of importance was attached to the survey by some of the respondents. The researcher was either greeted with reluctance or nonchalance. In some cases contradictory answerers were also given which seemed to show that the respondents were trying to get over the interview as soon as possible In many cases it was also felt that when a respondent was confused as to a particular response he/she usually responded with the same answer that the researcher hinted at.

The researcher has to be careful, when making use of secondary data\(^4\), of the definitions used by those responsible for its preparation, for example, researchers are interested in rural communities and their average family size. If published statistics are consulted

\(^3\) Uwe, F. (2011). Introducing research methodology: A beginner’s guide to doing a research project, United Kingdom, 13-40.

then a check must be done on how terms such as “family size” have been defined. They may refer only to the nucleus family or include the extended family. Even apparently simple terms such as ‘farm size’ need careful handling. Such figures may refer to any one of the following: the land an individual owns, the land an individual owns plus any additional land he/she rents, the land an individual owns minus any land he/she rents out, all of his land or only that part of it which he actually cultivates. It should be noted that definitions may change over time and where this is not recognized erroneous conclusions may be drawn.

Geographical areas may have their boundaries redefined, units of measurement and grades may change and imported goods can be reclassified from time to time for purposes of levying customs and excise duties. When a researcher conducts fieldwork she/he is possibly able to estimate inaccuracies in measurement through the standard deviation and standard error, but these are sometimes not published in secondary sources. The only solution is to try to speak to the individuals involved in the collection of the data to obtain some guidance on the level of accuracy of the data. The problem is sometimes not so much ‘error’ but differences in levels of accuracy required by decision makers. When the research has to do with large investments in, say, food manufacturing, management will want to set very tight margins of error in making market demand estimates. In other cases, having a high level of accuracy is not so critical. For instance, if a food manufacturer is merely assessing the prospects for one more flavour for a snack food already produced by the company then there is no need for highly accurate estimates in order to make the investment decision.
Researchers have to be aware of vested interests when they consult secondary sources. Those responsible for their compilation may have reasons for wishing to present a more optimistic or pessimistic set of results for their organization. It is not unknown, for example, for officials responsible for estimating food shortages to exaggerate figures before sending aid requests to potential donors. Similarly, and with equal frequency, commercial organisations have been known to inflate estimates of their market shares.

The reliability of published statistics may vary over time. It is not uncommon, for example, for the systems of collecting data to have changed over time but without any indication of this to the reader of published statistics. Geographical or administrative boundaries may be changed by government, or the basis for stratifying a sample may have altered. Other aspects of research methodology that affect the reliability of secondary data are the sample size, response rate, questionnaire design and modes of analysis.

Most censuses take place at 10 year intervals, so data from this and other published sources may be out-of-date at the time the researcher wants to make use of the statistics.

The time period during which secondary data was first compiled may have a substantial effect upon the nature of the data. For instance, the significant increase in the price obtained for Ugandan coffee in the mid-90’s could be interpreted as evidence of the effectiveness of the rehabilitation programme that set out to restore coffee estates which had fallen into a state of disrepair. However, more knowledgeable coffee market experts

---

would interpret the rise in Ugandan coffee prices in the context of large scale destruction of the Brazilian coffee crop, due to heavy frosts, in 1994, Brazil being the largest coffee producer in the world.

Whenever possible, marketing researchers ought to use multiple sources of secondary data. In this way, these different sources can be cross-checked as confirmation of one another. Where differences occur an explanation for these must be found or the data should be set aside.

The researcher faced problems while collecting any type of data related to DMRC as it was only available on the DMRC’s website: http://www.delhimetrorail.com and few books published by DMRC. There was no literature review available for service quality on DMRC. The records available on the website of DMRC were only updated till year 2011. There was paucity of literature available on DMRC. There was a scarcity of service quality research papers of Metro or subway trains therefore the literature was reviewed on public transport.

The research gave a first hand experience of understanding the mentality of commuters.

The present chapter contemplates on the relationship between the demographic variables namely, age and gender of the commuters and Metro usage.

Demographics refer to the statistical data of a population, especially those showing average age, income, education. For this study out of all the demographic factors mainly two age and gender have been focused.

---

5.2. DESCRIPTIVE STATISTICS

Table 5.1: Descriptive statistics

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of total items</td>
<td>1200</td>
</tr>
<tr>
<td>2</td>
<td>Number of incomplete forms</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>(1200-120) n=1080</td>
</tr>
</tbody>
</table>

Source: SPSS output tables

Table 5.1 depicts that out of a total of 1200 questionnaires, 120 were judged unusable due to missing responses, leaving a final sample size of n = 1080.

5.3. DESCRIPTIVE STATISTICS FOR DEMOGRAPHY (AGE AND GENDER)

Table 5.2: Demographic summary (Age & Gender) (Refer to Appendix Table 1 and 2)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Demographic variable</th>
<th>Testing variable</th>
<th>Valid case</th>
<th>Missing case</th>
<th>Percentage</th>
<th>Total case</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td>Perception of metro usage</td>
<td>1080</td>
<td>0</td>
<td>Males 51.7% Females 48.3%</td>
<td>1080</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Perception of metro usage</td>
<td>1080</td>
<td>0</td>
<td>18-29 38.3% 30-39 50.8% 40 Above 10.8%</td>
<td>1080</td>
</tr>
</tbody>
</table>

Source: SPSS output tables

In table 5.2 descriptive statistics we try to describe the demographics of the sample. In this case two demographic variables namely, Gender and Age have been studied; the testing variable which has been used is the perception of metro usage. Since the data was normally distributed parametric tests have been conducted.

Table 5.2 shows that the total numbers of valid cases are 1080 and there is no missing value in the data.
Gender

Table 5.2 depicts that, for gender out of the data collected from 1080 respondents:

- 51.7% i.e. 558 respondents were male
- 48.3% i.e. 522 respondents were female.

Age

Table 5.2 shows that for Age, out of the data collected from 1080 respondents

- 38.3% i.e. 414 respondents were from the age group of 18-29
- 50.8% i.e. 549 respondents were from the age group of 30-39
- 10.8% 117 respondents were above 40.

5.3.1. Gender - Graphical Presentation

Source: SPSS output tables

Figure 5.1: Gender
Figure 5.1 shows that out of the data collected from 1080 respondents:

- 51.7% i.e. 558 respondents were male
- 48.3% i.e. 522 respondents were female

5.3.2. Age Graphical Representation

![Age Graphical Representation](image)

Source: SPSS output tables

**Figure 5.2: Age**

Figure 5.2 shows that out of the data collected from 1080 respondents:

- 38.3% that is 414 respondents were from the age group of 18-29
- 50.8% i.e. 549 respondents were from the age group of 30-39
- 10.8% 117 respondents were above 40.
5.4. TEST OF NORMALITY

Test of normality is conducted to see whether the sample drawn is normally distributed or not. If the data is normally distributed we conducted parametric tests.

First, normality of each item was assessed using standard deviation, skewness, and kurtosis. There was no item that exhibited abnormally high standard deviation, skewness, and kurtosis, indicating normal distribution of each item.

Table 5.3: Table of skewness

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Valid</td>
<td>1080</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Mean</td>
<td>3.53</td>
</tr>
<tr>
<td>3</td>
<td>Std. Deviation</td>
<td>1.169</td>
</tr>
<tr>
<td>4</td>
<td>Skewness</td>
<td>-.264</td>
</tr>
<tr>
<td>5</td>
<td>Std. Error of Skewness</td>
<td>.074</td>
</tr>
<tr>
<td>6</td>
<td>Kurtosis</td>
<td>-.913</td>
</tr>
<tr>
<td>7</td>
<td>Std. Error of Kurtosis</td>
<td>.149</td>
</tr>
</tbody>
</table>

Source: SPSS output tables

Table 5.3 shows that the total number of items is 1080 for which the mean is 3.53 and the standard deviation is 1.169. And it can also be clearly seen that the data is normally distributed since the value of skewness and kurtosis lies between +1 and -1 that is -0.264 and -0.913 respectively.
Figure 5.3 shows that the number of items in the data is 1080 for which the mean is 3.53 and the standard deviation is 1.169.

The bell shaped frequency curve also depicts that the data is normally distributed.

5.5. SUMMARY OF CHI-SQUARE

The Demographics (Age and Gender)

Chi square is conducted when the data is in cross Tabulation (Crosstabs). Cross tabulation is a statistical process through which categorical data is converted in to contingency table.
Table 5.4: Summary of chi square

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Independent variable</th>
<th>Testing variable</th>
<th>Pearson Chi-square significant value</th>
<th>Contingency coefficient</th>
<th>Lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td>Perception of metro usage</td>
<td>.000</td>
<td>.221</td>
<td>.097</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Perception of metro usage</td>
<td>.000</td>
<td>.422</td>
<td>.214</td>
</tr>
</tbody>
</table>

Source: SPSS output tables

5.5.1. Gender

The cross-tabulation table produced by SPSS tried to find a relationship between the gender of the respondents (independent variable) and the overall perception of usage of Delhi Metro (dependent variable). The reason for gender being independent variable is because it has been found generally that gender of the respondent determines the overall perception of usage of Delhi Metro (Refer to appendix table 3).

Table 5.4 shows that there is a significant relationship between gender of the respondent and overall perception of usage of Delhi Metro. The significance can be proved statistically with the help of Pearson’s chi-square.

From the table (Refer Appendix Table 3a), it has been found that the significant value is 0.000 which is less than 0.05 at 95% confidence level. But as the thumb rule, the significant value has to be less than 0.05 at 95% confidence level. In this case, the small value of Pearson’s Chi-square test states that there is a significant interrelationship between gender and overall perception of usage of Delhi Metro.

So at 95% confidence level 100-95=5 divided by 100 or 0.05 significant level, it is concluded that there is a significant interrelationship between overall perception of usage of Delhi Metro and gender.
Lambda is a measure of reduction in error in measuring the association between the two variables. For example if the value of Lambda is 0.09, it implies that it is leading to a 9% reduction in error in estimating or predicting one variable from the other. In this case the value of Lambda is .097 (refer Appendix table 3b), which means that there is 9.7 percent error reduction. This is quite a small value so it can be concluded that there is a moderate relationship between the two variables, but statistically significant.

The contingency coefficient gives the measure of strength of the output. If the value is close to 1, there is strong correlation between the two variables. However, if the range is between 0.5 and 1, there exist a strong correlation, it can be concluded that there is moderate correlation between the variables namely Gender and perception overall perception of usage of Delhi Metro since the value is .221 (refer Appendix Table 3c), which is significant at 95% confidence level since the value of significance is 0.00.

5.5.2. Age

The cross-tabulation table (refer Appendix table 4) produced by SPSS tries to find a relationship between the Age of the respondents (independent variable) and the overall perception of usage of Delhi Metro (dependent Variable). The reason for age being independent variable is because it has been found generally that age of the respondent determines the overall perception of Delhi Metro.

From the table, (refer Appendix table 4a) it has been found that the significant value is 0.000 which is less than 0.05 at 95% confidence level. But as the thumb rule the significant value has to be less than 0.05 at 95% confidence level. In this case, the small value of Pearson’s Chi-square test states that there is a significant interrelationship between age and overall perception of usage of Delhi Metro.
So at 95% confidence level $100-95=5$ divided by 100 or 0.05 significant level, it is concluded that there is a significant interrelationship between overall perception of usage of Delhi Metro and age.

Lambda is a measure of reduction in error in measuring the association between the two variables. For example if the value of Lambda is 0.2, it implies that it is leading to a 20% reduction in error in estimating or predicting one variable from the other. The table above shows that the value of Lambda is 0.214 (refer Appendix table 4b), which means that there is 21.40 percent error reduction. This is quite a small value so it can be concluded that there is a moderate relationship between the two variables, but statistically significant.

The contingency coefficient gives the measure of strength of the output. If the value is close to 1, there is strong correlation between the two variables. However, if the range is between 0.5 and 1, there exists a strong correlation. From the table above, since the value of contingency coefficient is 0.422 (refer Appendix table 4c). It can concluded that there is moderate correlation between the variables namely age and overall perception of usage of Delhi Metro.