Project Report Part 1

IE 3301-004

Muhammad Muawiz Farooqi

# Introduction

The main goal of this project is to gather and then analyze real-world data. This project involves sampling two sets of data from the real-world and summarizing each set of data statistically and presenting the process and findings in this report.

# Data Collection

Data set 1 was obtained from [Kaggle](https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByMajorCity.csv). This dataset contains the surface temperatures for most major cities between the 1900s and 2013. This project analyses the data for Madrid, Spain between 2001 and 2013. This data was compiled using the Berkeley Earth website. It could have been collected from laboratory stations once each month.

Data set 2 was obtained from [Kaggle](https://www.kaggle.com/datasets/mysarahmadbhat/nyc-traffic-accidents). This dataset contains dates and times for all New York City Traffic accidents between January and August 2020. This project analyses the data for crashes in all boroughs of NYC on March 1, 2020. This data was obtained from reports of motor vehicle collisions by the New York City Police Department (NYPD).

# Descriptive Statistics

The analysis for both sets of data was done in Microsoft Excel.

**Data Set 1**

**Chart, box and whisker chart

Description automatically generated**

|  |  |
| --- | --- |
|  | Temperature (°C) |
| Sample mean | 12.34 |
| Sample standard deviation | 6.75 |
| Q1 | 6.32 |
| Q2 | 11.42 |
| Q3 | 18.46 |

Descriptive statistics for set 1

Frequency table for set 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lower Limit (°C) | Upper Limit (°C) | Frequency | Relative Frequency | Cumulative Relative Frequency |
| 1.746 | 3.506 | 7 | 0.047 | 0.047 |
| 3.506 | 5.266 | 23 | 0.153 | 0.2 |
| 5.266 | 7.026 | 15 | 0.1 | 0.3 |
| 7.026 | 8.786 | 12 | 0.08 | 0.38 |
| 8.786 | 10.546 | 13 | 0.087 | 0.467 |
| 10.546 | 12.306 | 10 | 0.067 | 0.534 |
| 12.306 | 14.066 | 11 | 0.073 | 0.607 |
| 14.066 | 15.826 | 6 | 0.04 | 0.647 |
| 15.826 | 17.586 | 9 | 0.06 | 0.707 |
| 17.586 | 19.346 | 10 | 0.067 | 0.774 |
| 19.346 | 21.106 | 11 | 0.073 | 0.847 |
| 21.106 | 22.866 | 16 | 0.107 | 0.954 |
| 22.866 | 24.626 | 7 | 0.047 | 1 |
|  |  | 0 | 1 |  |

Chart, histogram

Description automatically generated

The average temperature in Madrid between 2001 and 2013 was 12.34 °C. The midpoint of the temperatures was 11.42 °C and the middle 50% of temperatures were all between 6.32 °C and 18.46 °C.

The temperatures were quite spread out, with a majority occurring at the higher and lower ends of the temperature range, highlighting the extremes between the cold winters and hot summers in the capital of Spain.

This set of data for the average surface temperature in Madrid between 2001 and 2013 does not appear to follow a normal distribution, although there could be a normal distribution if the temperatures in the summer and winter are graphed separately.

**Data Set 2**

Chart, box and whisker chart

Description automatically generated

|  |  |
| --- | --- |
|  | Time (h: mm) |
| Sample mean | 0:09 |
| Sample standard deviation | 0:09 |
| Q1 | 0:03 |
| Q2 | 0:07 |
| Q3 | 0:12 |

Descriptive statistics for set 2

Frequency table for set 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lower Limit (h: mm) | Upper Limit (h: mm) | Frequency | Relative Frequency | Cumulative Relative Frequency |
| 0:00 | 0:04 | 40 | 0.276 | 0.276 |
| 0:04 | 0:08 | 33 | 0.228 | 0.504 |
| 0:08 | 0:12 | 35 | 0.241 | 0.745 |
| 0:12 | 0:16 | 14 | 0.097 | 0.842 |
| 0:16 | 0:20 | 4 | 0.028 | 0.87 |
| 0:20 | 0:24 | 5 | 0.034 | 0.904 |
| 0:24 | 0:28 | 6 | 0.041 | 0.945 |
| 0:28 | 0:32 | 1 | 0.007 | 0.952 |
| 0:32 | 0:36 | 3 | 0.021 | 0.973 |
| 0:36 | 0:40 | 0 | 0 | 0.973 |
| 0:40 | 0:44 | 3 | 0.021 | 0.994 |
| 0:44 | 0:48 | 1 | 0.007 | 1 |
|  |  | 145 | 1 |  |

Chart, histogram

Description automatically generated

The average time between motor vehicle crashes in New York City on March 1, 2020, was 9 minutes. The midpoint of the time between crashes was 7 minutes and the middle 50% of times between accidents were all between 3 and 12 minutes.

The times between accidents were skewed to left, with a large majority of accidents occurring within 12 minutes apart from one another, highlighting the frequent occurrences of motor vehicle accidents in a single day in New York City.

This set of data for the average time between motor vehicle crashes in New York City on March 1, 2020, appears to follow an exponential distribution, as shorter lengths of time between accidents are more frequent, and longer lengths are less frequent.

# Conclusion

Two sets of real-world data were obtained from the Kaggle website and were analyzed to observe trends. The first data set was about earth surface temperatures and did not follow a normal distribution as previously expected but may contain underlying normal distribution curves. The second data set for times between vehicle accidents followed an exponential distribution and was skewed to the left.

# Appendices

Appendix 1

|  |  |
| --- | --- |
| **Date** | **Average Temperature** |
| 1/1/2001 | 5.441 |
| 2/1/2001 | 5.755 |
| 3/1/2001 | 10.06 |
| 4/1/2001 | 10.074 |
| 5/1/2001 | 14.17 |
| 6/1/2001 | 20.379 |
| 7/1/2001 | 21.01 |
| 8/1/2001 | 22.62 |
| 9/1/2001 | 17.434 |
| 10/1/2001 | 13.859 |
| 11/1/2001 | 5.409 |
| 12/1/2001 | 1.746 |
| 1/1/2002 | 5.265 |
| 2/1/2002 | 6.35 |
| 3/1/2002 | 8.581 |
| 4/1/2002 | 10.432 |
| 5/1/2002 | 12.783 |
| 6/1/2002 | 20.186 |
| 7/1/2002 | 21.523 |
| 8/1/2002 | 20.655 |
| 9/1/2002 | 17.008 |
| 10/1/2002 | 12.711 |
| 11/1/2002 | 8.125 |
| 12/1/2002 | 6.571 |
| 1/1/2003 | 3.564 |
| 2/1/2003 | 3.927 |
| 3/1/2003 | 9.624 |
| 4/1/2003 | 10.377 |
| 5/1/2003 | 15.1 |
| 6/1/2003 | 22.305 |
| 7/1/2003 | 22.528 |
| 8/1/2003 | 24.524 |
| 9/1/2003 | 18.952 |
| 10/1/2003 | 11.325 |
| 11/1/2003 | 8.229 |
| 12/1/2003 | 4.634 |
| 1/1/2004 | 5.265 |
| 2/1/2004 | 5.253 |
| 3/1/2004 | 6.593 |
| 4/1/2004 | 8.991 |
| 5/1/2004 | 12.617 |
| 6/1/2004 | 21.125 |
| 7/1/2004 | 22.177 |
| 8/1/2004 | 21.352 |
| 9/1/2004 | 19.14 |
| 10/1/2004 | 13.078 |
| 11/1/2004 | 6.235 |
| 12/1/2004 | 4.243 |
| 1/1/2005 | 2.646 |
| 2/1/2005 | 2.125 |
| 3/1/2005 | 7.878 |
| 4/1/2005 | 10.763 |
| 5/1/2005 | 15.782 |
| 6/1/2005 | 21.884 |
| 7/1/2005 | 22.923 |
| 8/1/2005 | 22.415 |
| 9/1/2005 | 17.637 |
| 10/1/2005 | 13.101 |
| 11/1/2005 | 6.315 |
| 12/1/2005 | 3.683 |
| 1/1/2006 | 2.656 |
| 2/1/2006 | 3.8 |
| 3/1/2006 | 8.566 |
| 4/1/2006 | 11.388 |
| 5/1/2006 | 16.416 |
| 6/1/2006 | 20.639 |
| 7/1/2006 | 24.298 |
| 8/1/2006 | 21.052 |
| 9/1/2006 | 19.334 |
| 10/1/2006 | 14.448 |
| 11/1/2006 | 10.054 |
| 12/1/2006 | 4.032 |
| 1/1/2007 | 4.315 |
| 2/1/2007 | 6.868 |
| 3/1/2007 | 7.149 |
| 4/1/2007 | 10.558 |
| 5/1/2007 | 13.661 |
| 6/1/2007 | 17.417 |
| 7/1/2007 | 21.366 |
| 8/1/2007 | 20.619 |
| 9/1/2007 | 18.215 |
| 10/1/2007 | 12.238 |
| 11/1/2007 | 6.443 |
| 12/1/2007 | 3.816 |
| 1/1/2008 | 5.708 |
| 2/1/2008 | 7.133 |
| 3/1/2008 | 7.585 |
| 4/1/2008 | 10.576 |
| 5/1/2008 | 12.996 |
| 6/1/2008 | 18.154 |
| 7/1/2008 | 21.296 |
| 8/1/2008 | 21.843 |
| 9/1/2008 | 16.876 |
| 10/1/2008 | 11.633 |
| 11/1/2008 | 5.491 |
| 12/1/2008 | 3.709 |
| 1/1/2009 | 3.005 |
| 2/1/2009 | 5.23 |
| 3/1/2009 | 9.026 |
| 4/1/2009 | 9.18 |
| 5/1/2009 | 15.945 |
| 6/1/2009 | 20.272 |
| 7/1/2009 | 22.57 |
| 8/1/2009 | 23.71 |
| 9/1/2009 | 18.501 |
| 10/1/2009 | 14.611 |
| 11/1/2009 | 9.194 |
| 12/1/2009 | 4.264 |
| 1/1/2010 | 3.443 |
| 2/1/2010 | 4.061 |
| 3/1/2010 | 6.672 |
| 4/1/2010 | 11.455 |
| 5/1/2010 | 12.979 |
| 6/1/2010 | 18.203 |
| 7/1/2010 | 23.974 |
| 8/1/2010 | 22.934 |
| 9/1/2010 | 18.323 |
| 10/1/2010 | 11.684 |
| 11/1/2010 | 6.327 |
| 12/1/2010 | 4.022 |
| 1/1/2011 | 4.232 |
| 2/1/2011 | 5.804 |
| 3/1/2011 | 7.579 |
| 4/1/2011 | 13.499 |
| 5/1/2011 | 16.44 |
| 6/1/2011 | 19.625 |
| 7/1/2011 | 21.49 |
| 8/1/2011 | 22.852 |
| 9/1/2011 | 19.865 |
| 10/1/2011 | 14.487 |
| 11/1/2011 | 8.841 |
| 12/1/2011 | 4.604 |
| 1/1/2012 | 3.753 |
| 2/1/2012 | 2.947 |
| 3/1/2012 | 8.808 |
| 4/1/2012 | 8.126 |
| 5/1/2012 | 15.943 |
| 6/1/2012 | 20.234 |
| 7/1/2012 | 21.831 |
| 8/1/2012 | 23.021 |
| 9/1/2012 | 18.204 |
| 10/1/2012 | 12.509 |
| 11/1/2012 | 7.61 |
| 12/1/2012 | 4.972 |
| 1/1/2013 | 4.691 |
| 2/1/2013 | 4.403 |
| 3/1/2013 | 7.504 |
| 4/1/2013 | 9.483 |
| 5/1/2013 | 11.629 |
| 6/1/2013 | 17.565 |

Set 1 raw data

Appendix 2

|  |  |
| --- | --- |
| **Crash Date** | **Crash Time** |
| 3/1/2020 | 12:00 AM |
| 3/1/2020 | 12:00 AM |
| 3/1/2020 | 12:05 AM |
| 3/1/2020 | 12:05 AM |
| 3/1/2020 | 12:30 AM |
| 3/1/2020 | 12:42 AM |
| 3/1/2020 | 1:28 AM |
| 3/1/2020 | 1:30 AM |
| 3/1/2020 | 1:50 AM |
| 3/1/2020 | 2:05 AM |
| 3/1/2020 | 2:29 AM |
| 3/1/2020 | 2:40 AM |
| 3/1/2020 | 2:45 AM |
| 3/1/2020 | 3:05 AM |
| 3/1/2020 | 3:15 AM |
| 3/1/2020 | 3:20 AM |
| 3/1/2020 | 3:46 AM |
| 3/1/2020 | 3:50 AM |
| 3/1/2020 | 3:50 AM |
| 3/1/2020 | 4:30 AM |
| 3/1/2020 | 4:40 AM |
| 3/1/2020 | 4:48 AM |
| 3/1/2020 | 5:00 AM |
| 3/1/2020 | 5:15 AM |
| 3/1/2020 | 5:25 AM |
| 3/1/2020 | 5:35 AM |
| 3/1/2020 | 5:50 AM |
| 3/1/2020 | 6:25 AM |
| 3/1/2020 | 6:30 AM |
| 3/1/2020 | 6:39 AM |
| 3/1/2020 | 6:50 AM |
| 3/1/2020 | 6:54 AM |
| 3/1/2020 | 7:35 AM |
| 3/1/2020 | 7:45 AM |
| 3/1/2020 | 8:20 AM |
| 3/1/2020 | 8:30 AM |
| 3/1/2020 | 8:30 AM |
| 3/1/2020 | 9:00 AM |
| 3/1/2020 | 9:10 AM |
| 3/1/2020 | 9:30 AM |
| 3/1/2020 | 9:40 AM |
| 3/1/2020 | 9:49 AM |
| 3/1/2020 | 9:50 AM |
| 3/1/2020 | 9:50 AM |
| 3/1/2020 | 10:00 AM |
| 3/1/2020 | 10:00 AM |
| 3/1/2020 | 10:00 AM |
| 3/1/2020 | 10:20 AM |
| 3/1/2020 | 10:40 AM |
| 3/1/2020 | 10:54 AM |
| 3/1/2020 | 11:00 AM |
| 3/1/2020 | 11:05 AM |
| 3/1/2020 | 11:29 AM |
| 3/1/2020 | 11:45 AM |
| 3/1/2020 | 11:58 AM |
| 3/1/2020 | 12:00 PM |
| 3/1/2020 | 12:00 PM |
| 3/1/2020 | 12:08 PM |
| 3/1/2020 | 12:22 PM |
| 3/1/2020 | 12:30 PM |
| 3/1/2020 | 12:30 PM |
| 3/1/2020 | 12:45 PM |
| 3/1/2020 | 12:50 PM |
| 3/1/2020 | 12:52 PM |
| 3/1/2020 | 12:56 PM |
| 3/1/2020 | 1:00 PM |
| 3/1/2020 | 1:10 PM |
| 3/1/2020 | 1:10 PM |
| 3/1/2020 | 1:20 PM |
| 3/1/2020 | 1:30 PM |
| 3/1/2020 | 1:41 PM |
| 3/1/2020 | 1:41 PM |
| 3/1/2020 | 1:45 PM |
| 3/1/2020 | 2:00 PM |
| 3/1/2020 | 2:15 PM |
| 3/1/2020 | 2:15 PM |
| 3/1/2020 | 2:17 PM |
| 3/1/2020 | 2:21 PM |
| 3/1/2020 | 2:23 PM |
| 3/1/2020 | 2:24 PM |
| 3/1/2020 | 2:26 PM |
| 3/1/2020 | 2:30 PM |
| 3/1/2020 | 2:33 PM |
| 3/1/2020 | 2:43 PM |
| 3/1/2020 | 2:45 PM |
| 3/1/2020 | 2:50 PM |
| 3/1/2020 | 2:54 PM |
| 3/1/2020 | 2:56 PM |
| 3/1/2020 | 3:00 PM |
| 3/1/2020 | 3:03 PM |
| 3/1/2020 | 3:05 PM |
| 3/1/2020 | 3:10 PM |
| 3/1/2020 | 3:18 PM |
| 3/1/2020 | 3:35 PM |
| 3/1/2020 | 3:53 PM |
| 3/1/2020 | 3:54 PM |
| 3/1/2020 | 4:00 PM |
| 3/1/2020 | 4:00 PM |
| 3/1/2020 | 4:09 PM |
| 3/1/2020 | 4:20 PM |
| 3/1/2020 | 4:25 PM |
| 3/1/2020 | 4:30 PM |
| 3/1/2020 | 4:30 PM |
| 3/1/2020 | 4:40 PM |
| 3/1/2020 | 4:43 PM |
| 3/1/2020 | 4:50 PM |
| 3/1/2020 | 4:50 PM |
| 3/1/2020 | 4:58 PM |
| 3/1/2020 | 5:00 PM |
| 3/1/2020 | 5:12 PM |
| 3/1/2020 | 5:15 PM |
| 3/1/2020 | 5:25 PM |
| 3/1/2020 | 5:29 PM |
| 3/1/2020 | 5:30 PM |
| 3/1/2020 | 5:34 PM |
| 3/1/2020 | 6:00 PM |
| 3/1/2020 | 6:02 PM |
| 3/1/2020 | 6:04 PM |
| 3/1/2020 | 6:10 PM |
| 3/1/2020 | 6:10 PM |
| 3/1/2020 | 6:11 PM |
| 3/1/2020 | 6:20 PM |
| 3/1/2020 | 6:30 PM |
| 3/1/2020 | 6:39 PM |
| 3/1/2020 | 6:46 PM |
| 3/1/2020 | 7:00 PM |
| 3/1/2020 | 7:08 PM |
| 3/1/2020 | 7:10 PM |
| 3/1/2020 | 7:15 PM |
| 3/1/2020 | 7:20 PM |
| 3/1/2020 | 7:30 PM |
| 3/1/2020 | 7:35 PM |
| 3/1/2020 | 7:53 PM |
| 3/1/2020 | 8:00 PM |
| 3/1/2020 | 8:10 PM |
| 3/1/2020 | 8:16 PM |
| 3/1/2020 | 8:30 PM |
| 3/1/2020 | 8:30 PM |
| 3/1/2020 | 8:40 PM |
| 3/1/2020 | 9:13 PM |
| 3/1/2020 | 9:20 PM |
| 3/1/2020 | 10:00 PM |
| 3/1/2020 | 10:24 PM |
| 3/1/2020 | 10:35 PM |
| 3/1/2020 | 10:40 PM |
| 3/1/2020 | 10:40 PM |

Set 2 raw data