Northumbria University
Student:
Aleksandar Peshev

# DATABASE AND ANALYTICS PRINCIPLES

Tableau Project

Northumbria University NEWCASTLE

| | | | |
|---|---|---|---|
| **Student** | Aleksandar Peshev | | |
| **Programme Title** | Computing and Technology MSc | | |
| **Module Title** | Database and Analytics Principles | | |
| **Module Code** | | | |
| **Module Tutor Name(s):** | | | |
| **Academic Year:** | 2021-2022 | | |
| **Coursework Title** | Database and Analytics Principles – Tableau Project | | |
| **Academic Declaration:**<br>*Yes, I am aware of Turnitin* | | | |
| **Word Count** | | **Date Submitted** | |

Northumbria University NEWCASTLE

# Table of Contents



## List of figures



# Business Intelligence – Tableau

## 1.0 Data Integrity Checks

Once opened and loaded on Tableau the data will be displayed as a data source tab and a sheet. While in the data source tab the data analyst should check for the data types. Which in this case are numbers for Orders ID, Quantity Order and Price Each, it is string for the product and purchase address and a date and time for the order date. Since having Order ID, Product, Quantity Order and Price Each it is safe to conclude that the Order's total field is missing. The next thing that needs checking is the description in each field. After checking all columns are valid which is good, some of them did not have their domain loaded and it had to be manually loaded. Lastly, in the tableau sheet it could be checked that Tableau divides the data columns in to two sections: all of the dimensions and all of the measures (Jurgon, 2021). It also needs to calculate the profit each, sales each and unit's price (all in Figure 12). A further brake down of the address is also needed which would be breaking down the orders by Address Breakdown (consisting of number and street), by City and by zip code. All thanks to Dreamer (2020) which also recommended a further break down of the date which is not really necessary for this task.

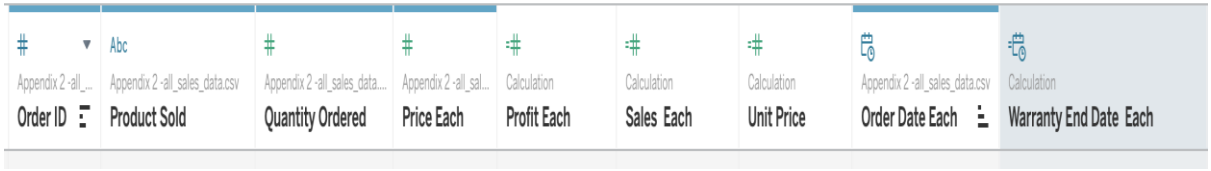


*Figure 1*

Figure 12: Data breakdown

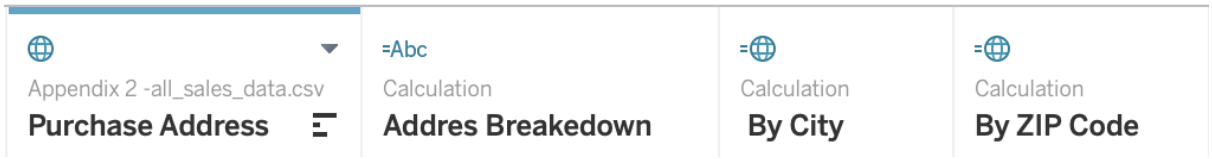


*Figure 2*

Figure 14: Address Breakdown

## 2.0 Warranty End Date calculation

DATEADD('month', 14, [Order Date])

By selecting DATEADD the analyst can add date in months and in case of this one the added time is 14 months (Figure 22) as a contingency (Jurgon, 2021).


*Figure 3*

Figure 22: Warranty End Date

## 3.0 Total Sales By City (highest to lowest)

Figure 15 below consist of the highest to lowest sales by city, which was done by dragging and dropping the necessary elements (sum of sales each and by city) and sorting them from highest to lowest (Tableau Step 4, 2021).

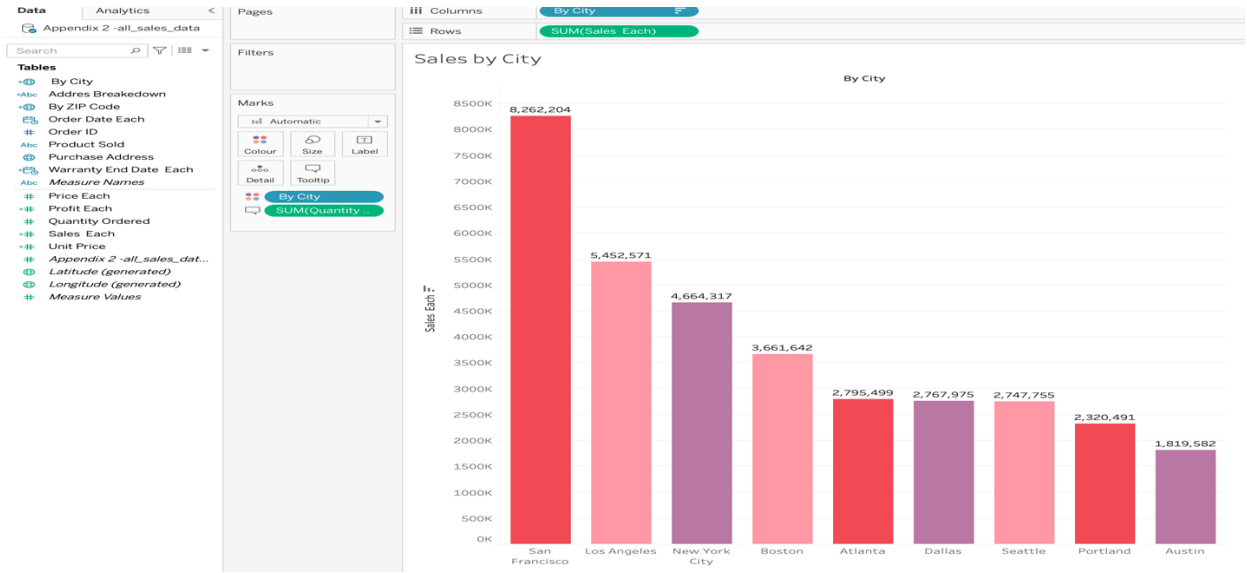


*Figure 4*

Figure 15

## 4.0 Cities generating more than 3 million US dollars for 2019

In Figure 16 are shown the four cities that made over 4 million US dollars 2019. This was achieved by dragging and dropping order date each and sales each into the columns section plus by city to the rows section. Furthermore, a filter needed adding in order to display the specific cities (Tableau Step 5, 2021).


*Figure 5*

Figure 16: Cities generating more that 3 million US dollars

## 5.0 Top selling Products

Here after using the product sold as Cullum and then putting order date each plus sales each in the rows section. Filters are added to the sales to display them as well as the order date which was limited to 2019 only. Last thing needed for this task was to add the quantity ordered through which tableau can be programed to show quantity sold in numbers, and the products sold where only the top 5 products would be selected manually (Tableau Step 5, 2021).

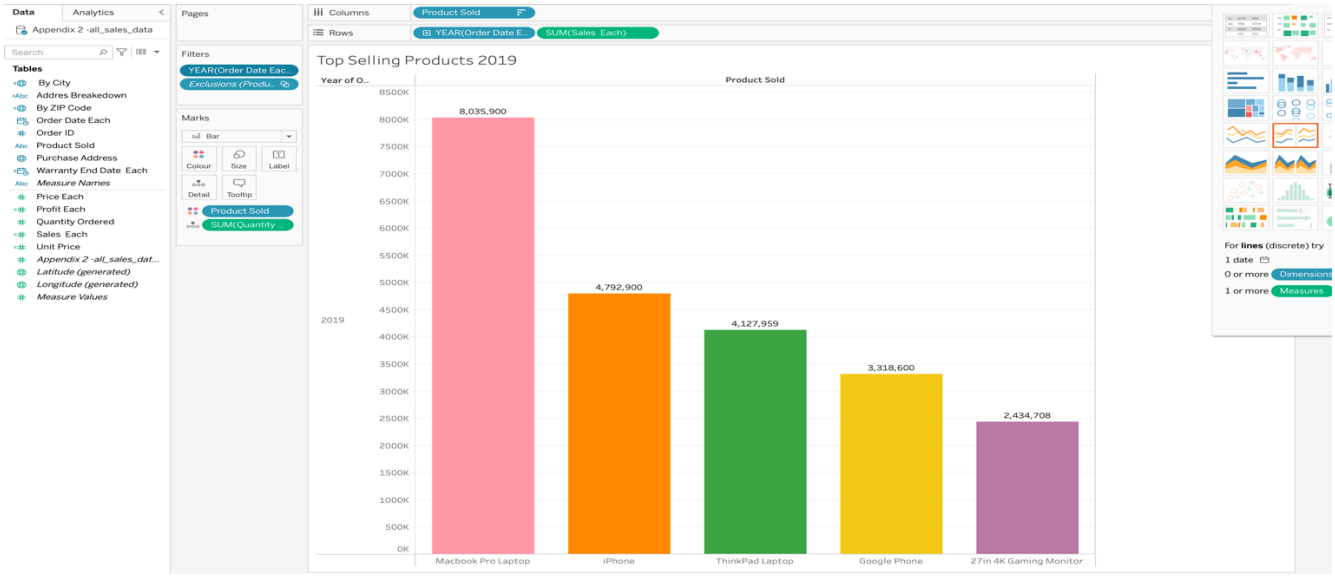


*Figure 6*

Figure 17:

## 6.0 Sales By month for 2019

What needs to be done in this part is for the rows to add the sales, and after that in the categories section the order date needs to be added three times as the first time the order date needs to be programmed for to show the year as well as in the filter, then the second one to show the quarter, the third for the month, as well as in the filter. Lastly, the brighter colours are to indicate increasing sales where darker colours show decreasing sales (Tableau Step 5, 2021).

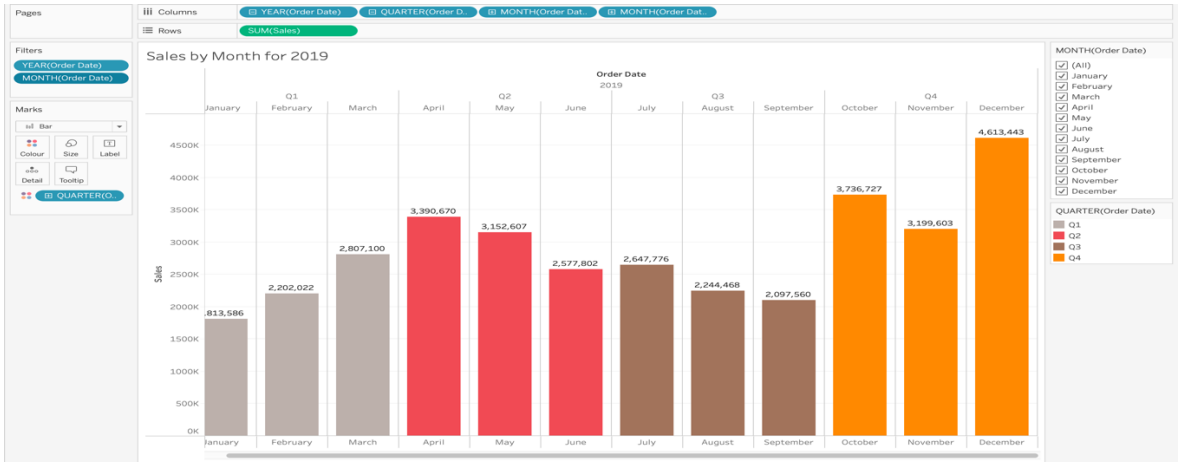


*Figure 7*

Figure 18: Sales By Month (2019)

## 7.0 Sales and Profit for 2019

In this part the columns need to include Order Date selected for year alongside the measure values. In the rows need to include the products and the measure names. In

the filter the year needs to be set for 2019, and the measure names have to be set for profit per order and the sales (Tableau Step 6 & 7, 2021).

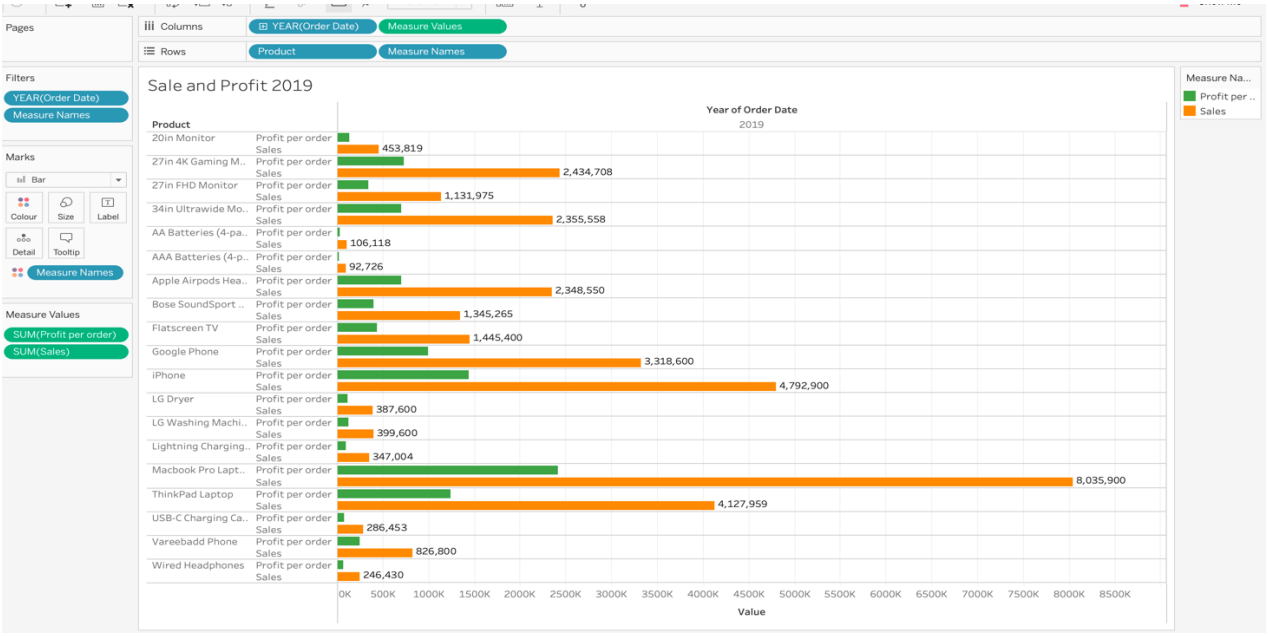


*Figure 8*

Figure 19: 2019 Sales and Profit

## 8.0 Maximum Sales for Boston, Los Angeles, Seattle and New York.

This part is fairly simple to achieve. Firstly, to display the cities need to be added (dragged and dropped) and filtered by manually selecting the selecting the four cities that are needed. Secondly, in the rows by adding the sales and then selecting to display the maximum sales (Tableau Step 5, 2021).

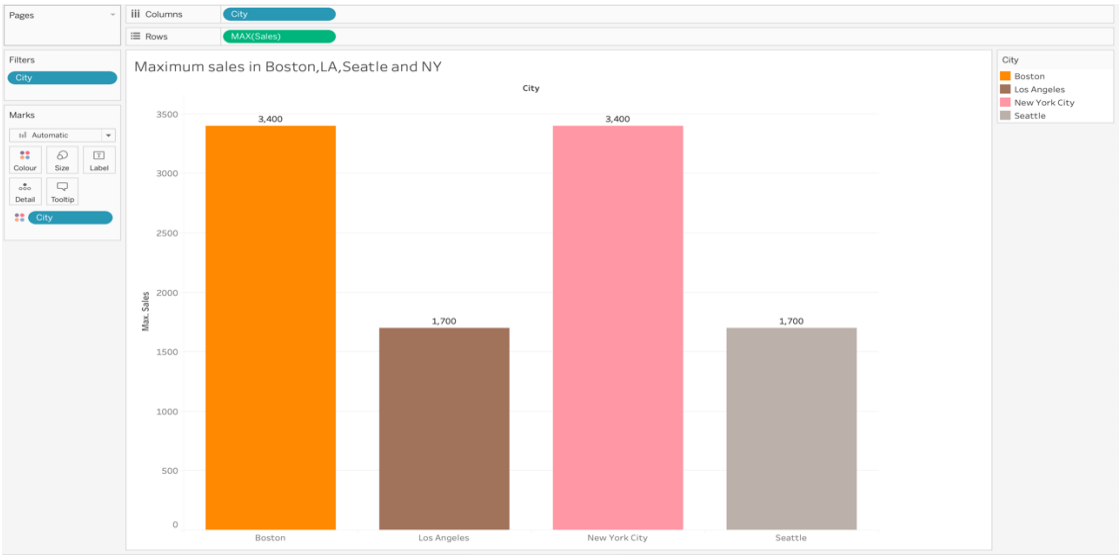


*Figure 9*

Figure 20: Maximum Sales

## 9.0 Dashboard for tasks 3,4,5 & 6.

What was done in this part was simply by selecting to create a new dashboard window and then from the sheets area dragging and dropping the sheets from tasks 3, 4, 5 & 6. Please note that I checked create a dashboard on the Tableau (2021) website, however this task was done by using my common sense and not guidance from Tableau's website.

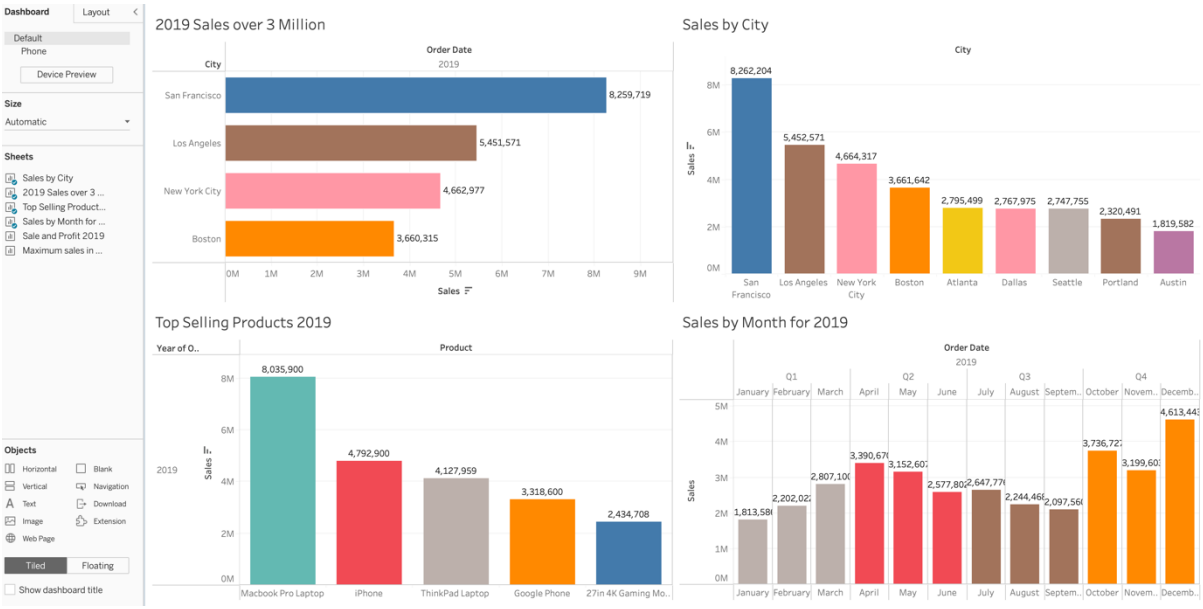


*Figure 10*

Figure 21: Dashboard for tasks 3-6

# References

Han, J., Kamber, M. & Pei, J. (2012) *Data Mining Concepts and Techniques*. 3rd edn. Elsevier Inc.: Waltham, MA, USA

Inmon, W.H. (2002) *Building the Data Warehouse*. 3rd edn. John Wiley & Sons, Inc.: Toronto, Canada

Kimball, R. & Ross, T. (2013) *The Data Warehouse Toolkit*. 3rd edn. John Wiley & Sons, Inc.: Indianapolis, IN, USA

Rob, P. & Coronel, C. (2009) *Database Systems: Design, Implementation, and Management*. 8th edn. Course Technology: Boston, MA, USA

Wang, J, (2003) *Data Mining: Opportunities and Challenges.* 1st edn. Idea Group Publishing: Hershey, PA, USA

Jatana, N., Puri, S., Ahuja, M., Kathuria, I. & Gosain, D. (2012) *A Survey And Comparison Of Relational And Non-Relational Base. International Journal Of Engineering Research & Technology*. 1 (6) pp. 1-5. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.9352&rep=rep1&type=pdf
(Accessed: 11/4/2021)

Niaksu, O. (2015) *CRISP Data Mining Methodology Extension For Medical Domain.* Baltic J. Modern Computing. 3 (2) pp. 92-109. Available at: https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/3_2_2_Niaksu.pdf
(Accessed: 24/4/2021)

Biscobing, J. (2021) *Data Sampling*. Available at:

https://searchbusinessanalytics.techtarget.com/definition/data-sampling (Accessed: 8/5/2021)

Becker, R. (2021) *Structured Query Language (SQL).* Available at: https://www.techopedia.com/definition/1245/structured-query-language-sql
Accessed: 16/3/2021)

Blue, J. (2012) *How To Use SQL in Statement With Subquery*. Available at: https://www.youtube.com/watch?v=ZmEpW-RVWxY (Accessed: 2/3/2021)

Cave Of Programming (2016) *My SQL Begginers 0006: Null and Not Null Values*. Available at:
https://www.youtube.com/watch?v=Ja4F4PcqrBc (Accessed; 3/3/2021)

CloudTweaks (2021) *When To Use Supervised and Unsupervised Data Learning*. Available at:
https://cloudtweaks.com/2014/09/supervised-unsupervised-data-mining/ (Accessed: 15/4/2021)

Dreamer, A. (20200 *How To Prepare Data For Tableau Guide: Step By Step.* Available at: https://www.xplenty.com/blog/xplenty-tableau-step-by-step/ (Accessed: 29/4/2021)

Guru99 (2021) *SQL vs NOSQL*. Available at: https://www.guru99.com/sql-vs-nosql.html (Accessed: 12/4/2021)

Hoare, J. (2021) *Machine Learning: Pruning Decision Trees*. Available at: https://www.displayr.com/machine-learning-pruning-decision-trees/ (Accessed: 8/5/2021)

Hans, S. (2019) *7 Reasons Data Preparation Is Important*. Available at: https://adeptia.com/blog/7-reasons-data-preparation-is-important (Accessed: 23/4/2021)

Jones, M. (2017) *Working With Messy Data*. Available at: https://developer.ibm.com/tutorials/ba-cleanse-process-visualize-data-set-1/ (Accessed: 9/5/2021)

Jurgon, P. (2021) *Data warehouse: Star Schema*. Available at: https://eu.bbcollab.com/collab/ui/session/playback (Accessed:

Jurgon, P. (2021) *Introduction To Tableau Part 3*. Available at: https://youtu.be/3PcMn3gZlDw (Accessed: 23/4/2021)

Jorgi, B. (2014) *SQL Server: DDL and DML Commands*. Available at: https://www.youtube.com/watch?v=Am1wIwKQWOo (Accessed: 1/3/2021)

Kudvenkat (2012) *Self Join In SQL server Part 14*. Available at: https://www.youtube.com/watch?v=qnYSN_7qwgg Accessed: 6/3/201)

Luthra, A. (2018) *Joins In SQL*. Available at: https://www.youtube.com/watch?v=M74w28kdbQg (Accessed: 5/3/2021)

LucidChart (2017) *Entity Relationship Diagram (ERD) Tutorial Part 1*. Available at: https://www.youtube.com/watch?v=QpdhBUYk7Kk ( Accessed: 16/2/2021)

MongoDB (2021) *Advantages of NOSQL*. Available at: https://www.mongodb.com/nosql-explained/advantages (Accessed: 17/3/2021)

MySQL Forum (2021) *MySQL Tutorials*. Available at: https://downloads.mysql.com/docs/mysql-tutorial-excerpt-5.7-en.pdf (Accesed: 5/4/2021)

Panwar, A. (2020) *DDL, DML And DCL In MySQL*. Available at: https://www.c-sharpcorner.com/UploadFile/65fc13/ddldml-and-dcl-in-mysql/ (accessed: 1/3/2021)

Poojary, D. (2021) *Type of NOSQL Databases and its Comparison To Relational Databases.* Available at: https://www.researchgate.net/profile/Dikshay-Poojary/publication/302557703_Article_Type_of_nosql_databases_and_its_comparison_with_relational_databases/links/5aeaa2b50f7e9b837d3c40e7/Article-Type-of-nosql-databases-and-its-comparison-with-relational-databases.pdf (17/3/2021)

QAFox (2020) *SQL Part 28: Gratest And Least SQL Functions.* Available at; https://www.youtube.com/watch?v=ftykz7QgL6A (accessed; 19/3/2021)

Solenz, D.L. (2021) *What are the advantages of Relational Data Model.* Available at: https://www.techwalla.com/articles/what-are-the-advantages-of-a-relational-database-model (Accessed: 1/4/2021)

Safe (2021) *What Is Data Conversion.* Available at: https://www.safe.com/what-is/data-conversion/ (Accessed: 9/5/2021)

TheNewBoston (2012) *MySQL Database Tutorial 19: Group By.* Available at: https://www.youtube.com/watch?v=_uyyc5fc3J8 (Accessed: 4/3/2021)

Tables Help (2021) *Step 4: Explore Your Data Geeoraphically.* Available at: https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-explore.htm (Accessed: 7/5/2021)

Tableau Help (2021) *Step 5: Drill Down The Detail*. Available at: https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-drilldown.htm (accessed: 7/5/2021)

Tableau Help (2021) *Step 6: Build A Dashboard To Show Your Insights.* Available at: https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-build.htm (Accessed: 9/5/2021)

Tableau Help (2021) *Step 7: Build A Story To Present.* Available At: https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-story.htm (Accessed: 9/5/2021)

Tableau (2021) *Create A Dashboard.* Available at: https://help.tableau.com/current/pro/desktop/en-us/dashboards_create.htm (Accessed: 9/5/2021)

Wirth, R. & Hipp, J. (No Date) *CRISP-DM: Towards A Standard Process Model For Data Mining*: Available at: http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf (Accessed: 24/4/2021)

Wakefield, K. (2021) *Predictive Analytics And Machine Learning.* Available at:

https://www.sas.com/en_gb/insights/articles/analytics/a-guide-to-predictive-analytics-and-machine-learning.html
(Accessed: 25/4/2021)

Wakefield, K. (2021) *A Guide To The Types Of Machine Learning Algorithms.* Available at:
https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html
(Accessed: 25/4/2021)