

Association for Information Systems

AIS Electronic Library (AISeL)

AMCIS 2022 Proceedings

SIG ED - IS in Education, IS Curriculum,
Education and Teaching Cases

Aug 10th, 12:00 AM

Data Analysis and Visualization with Google Data Studio

Christopher Njunge

California Lutheran University, chris.njunge@gmail.com

Follow this and additional works at: <https://aisel.aisnet.org/amcis2022>

Recommended Citation

Njunge, Christopher, "Data Analysis and Visualization with Google Data Studio" (2022). *AMCIS 2022 Proceedings*. 23.

https://aisel.aisnet.org/amcis2022/sig_ed/sig_ed/23

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Data Analysis and Visualization with Google Data Studio

Emergent Research Forum (ERF)

Christopher Njunge
California Lutheran University
cnjunge@callutheran.edu

Abstract

As part of an Information Systems course a project was created to illustrate how Google Data Studio can play a role in improving the analysis and visualization of data. Often practitioners interact with data from already existing databases or datasets and need to make sense of them before analyzing, visualizing and gleaning insights from them. The dataset used in this case is retrieved from the w3schools learn SQL online course. A query is written to extract the data, and the data is imported to Google Sheets for visualization with Google Data studio. The process enables the identification of data anomalies, the analysis of trends, visualization of outliers and creation of additional data to gain further insights. This process reduces the time of analysis and provides students with insights on the life cycle of data analysis from business understanding, to data understanding and back to business understanding. Further, results that may be expected such as performance of employees, products or vendors is reviewed in more detail to gain further insights.

Keywords

Data Visualization, Analytics, Data Studio.

Introduction

As the amount of data that organizations and society has to deal with grows rapidly, the requisite skills to analyze, visualize and interpret the data become more valuable. The number of companies offering products, solutions and courses related to data visualization has also grown in recent years. According to Gartner's Magic Quadrant for Analytics and Business Intelligence platforms, Microsoft, Salesforce (Tableau) and Qlick are the Leaders in the industry. (Gartner, 2022). However, Google and Domo are seen to be challengers, with Google's Cloud Platform ecosystem being noted for robust application development resources and sound architecture and governance. This development, the fact that it has basic functionality for data visualization and is free make for compelling reasons to introduce students to Google Data Studio. Use of such a free tool can serve to remove a barrier to data visualization techniques and Information Systems in general for students who may not have the resources or devices to access more mature and license based systems.

Project Overview

This project is offered as part of an undergraduate Information Systems course in a Liberal Arts College. There are no pre-requisites for the course and the project covers three chapters in the course, namely: Storing and Organizing Information, Analyzing Information for Business Decision Making and Information and Knowledge for Business Decision making. These chapters correspond with the chapters in the textbook used in the course. (Belanger, 2015)

The first part of the project involved the students reviewing the online database that is part of W3Schools 'Learn SQL' module. This database was selected because it provides a simple example

of a relational database. Students view the entities that make up the database and can click on them to view the attributes that make up the entity. In reviewing the attributes, the students can identify the existence of foreign keys in various entities and the idea of an Entity Relationship Diagram (ERD) is discussed. Students are then guided to create a full ERD of the database.

The second part of the project involves the students extracting the data from the database. A SQL statement is written to retrieve the data based on the relationships that exist between the entities. Writing SQL queries is not expressly covered in the project but students do see the syntax and the resulting table which is downloaded and saved as a Google Sheets document. The Google Sheets document serves as the dataset for the visualization using Google Data Studio.

Overall the project can fit well as part of an introductory course in a data visualization or analytics series of course in an Information Systems or related program.

Project Questions

This project seeks to answer the following questions based on the dataset.

1. Who is the top employee?
2. What is the top product?
3. Where is the top market?
4. Who is the top shipper?
5. What is the 'best' category?
6. What is a metric the company should be looking at that's not listed above?

Methods

Google Data Studio is a data visualization application that lets you visualize data from a variety of sources. Before using Google Data Studio (Google, n.d.). You need to have a 'Data Set' available that Google Data Studio can connect to and visualize. In this case W3store database stored in Google is used.

The first thing students see when they navigate to the Data Studio site is the option to 'Use it for Free'. Clicking on 'Use it For Free' gets you to the login page. Once prompted, they use their Gmail login credentials. This will allow them to access the data from their Google Drive. Once logged in they see the option to create a 'Blank Report'.

Clicking on 'Blank Report' creates a blank report that will bring up the next step of connecting to a dataset. The following instructions are provided to be able to import the dataset to Google Data Studio.

Step 1: Connect your data -

Click on 'Create New Data Source'.

Once you click 'Create New Data Source' you will be prompted to 'Select' your data source. In our example, we will be using 'Google Sheets'. Select 'Google Sheets'.

Your data set (W3SchoolsStore) should appear on the list of Spreadsheets on Google Sheets. Select it and accept the default of using the 'first row as headers'. Then click 'Connect' on the top right corner of the page.

Step 2: Create your first chart/visualization -

Once your data is connected you are ready to start creating charts/visualizations. Click on 'Insert' and select one of the options, in this example 'Bar Chart'. A 'plus' sign will appear after your selection, click the 'plus' sign into the empty workspace. A chart will appear with some default settings.

Step 3: Customize your chart/visualization Once you create the chart. A navigation panel will appear on the right side of the page. This panel will enable you to edit the chart to get it to display the data you intend to show. The panel includes a few key sections. To start edit 'Dimension' and 'Metric'. Clicking on these will show you options for the Dimensions and Metrics available. Select the ones that are most relevant for your

example. For this example, we use 'level 2' for Dimension, as this shows the groupings of ages in the data set, and 'value' for the number of residents that fall within this group.

Step 4: Focusing and filtering your data

To filter your data, select 'Insert' then select 'Filter Control'. Clicking on the filter will let you edit the filter on the panel on the right hand side. To view the chart with the filter in use, select 'View' on the top right part of the page.

Step 5: Creating a dashboard To create a dashboard. Go through the same steps above and add different types of charts that can answer questions you have in mind. You can add an additional data source to view data from different sources side by side.

Step 6: Exploring Geographic Data

To view Geographic data your data source must have geographic data e.g. Zip Codes, State or City Names etc. For this exercise we will need to edit the W3 store data to ensure that City and Country are stored as 'Geographic' information data types and not the default, text.

Go through the same steps as previously described to add the new Data Source. Once you insert the chart, navigate to the panel on the right hand side and update the 'Data Source' to the new Spreadsheet 'Superstore Data'.

Select 'Insert' and select 'Geo Map' Update the dimensions to 'City' and select 'Sales' for the metric to visualize.

Step 7: Summarize your findings To summarize your findings makes sure that 'Titles' are added to all your charts and that the titles describe what the charts are showing. For example, from the first chart we could use a title of 'Singapore Population by Age Group', and for the second chart 'Superstore Sales by U.S. City'. To add a title, select 'Insert' and then select 'Text'.

Step 8: Share your findings

To share you findings you can either:

- * 1) Save a copy of your report or dashboard by selecting 'File' then 'Download'. There are several options to download as including PDF or Webpage.
- * 2) Click on 'Share' on the top right corner and either invite people or get a 'Shareable Link'.

Results

A dashboard was developed to answer the questions proposed as part of the project and yielded the following results.

1. Who is the top employee?

The 'Summary of Employee Orders and Sales' chart showed that the top employee is Margaret based on both the number of orders and the sum of sales. Based on the number of orders the second placed employee was Janet but based on Sales, it was Nancey.

2. What is the top product?

The 'Summary of Product Orders and Sales' chart showed that the top product by sales was "Cte de Blaye" with more than \$60,000 in sales. There were three products tied for the most orders, 14: "Gorgonzola Telino, Mozzarella di Giovanni and Raclette Courdavault".

3. Where is the top market?

The 'Heat Map of Sales By City' chart showed that the top markets for sales were in Europe. Paris and Frankfurt had the highest number of sales.

4. Who is the top shipper?

The 'Summary of Orders by Shipper and Category' chart showed that the shipper with most shipments was United Package and the category with the most orders was Dairy.

5. What is the 'best' category?

The 'Summary of Orders by Shipper and Category' chart showed that the category with the most orders was Dairy.

6. What is a metric the company should be looking at that's not listed above?

Students are encouraged to come up with questions about the data that they can explore further such questions may require creation of new fields e.g. identifying the month or day with the most sales or identifying repeat or high value customers.

Discussion

This study showed that Google Data studio can be useful in quickly visualizing data and can be used to illustrate the iterative cycle of analyzing data and visualizing it. Two issues that came up that highlighted the iterative nature of analyzing data for visualization were incorrect city names and updating geographic data. Three cities had the incorrect name and were thus not identifiable as geographic information. This was identified using the 'Data Cleanup' function in Google Sheets. The correct name was identified by validating the street address and country for the relevant record. The city and country fields were updated in Data Studio from 'Text' fields to 'Geo' fields to enable their use in maps.

Conclusion

This project has proved to be an integral part of the Information Systems and Organizational Design course that it is a part of. Only anecdotal feedback has been collected from students, including what is provided in course evaluations. Some of the feedback has been very encouraging for example one student sought to apply this process of analyzing and visualizing data to their pitch in a startup competition and progressed through several levels of the competition. Another student sought to implement similar visualizations in their family owned restaurant. While the feedback is not conclusive it provides some insight that the students are engaging with the content and finding meaningful application in real world settings.

8. ACKNOWLEDGEMENTS

I would like to acknowledge and thank the reviewers and conference chairs for their constructive feedback and support.

REFERENCES

- Belanger, F., Van Slyke, C. & Crossler, R. E. *Information Systems for Business: An Experiential Approach* 2nd Edition, Wiley, 2015.
- Casner, S.M.: A task-analytic approach to the automated design of graphic presentation. *ACM Transactions on Graphics* 10, 111–151 (1991)
- Cleveland, W.S.: *Visualizing Data*. Hobart Press (1993)
- Dastani, M.: The Role of Visual Perception in DataVisualization. *Journal of Visual Languages and Computing* 13, 601–622 (2002)
- Gartner. (n.d.). *Products In Analytics and Business Intelligence Platforms Market* <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>
- Google. (n.d.). *Google Data studio overview*. Retrieved March 8, 2022, from <https://datastudio.google.com/>
- Jafar, M., Babb, J., and Abdullat, A. (2016). *Emergence of Data Analytics in the Information Systems Curriculum*, Proceedings of the EDSIG Conference, Las Vegas, NV.
- Johnson, C.R.: *Top Scientific Visualization Research Problems*. *IEEE Computer Graphics & Applications* 24, 13–17 (2004)

- Johnson, C., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P., Yoo, T.S.: NIH/NSF Visualization Research Challenges Report. IEEE Press (2006)
- Kosslyn, S.M.: Understanding Charts and Graphs. *Applied Cognitive Psychology* 3, 185–226 (1989)
- Primich, T. (2010). A Semester-Long Seminar in Statistical Visualization for Undergraduates as Taught by a Science and Engineering Librarian. *Science & Technology Libraries*, 29(3), 181-188.
- Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: Proceedings of the AVI workshop on Beyond time and errors: novel evaluation methods for information visualization, ACM, New York (2006)
- Tufte, E.R.: *The Visual Display of Quantitative Information*, 2nd edn. Graphics Press (2001)
- Vekiri, I.: What Is the Value of Graphical Displays in Learning? *Educational Psychology Review* 14, 261–312 (2002)
- Wijk, J.J.v.: The Value of Visualization. In: Proceedings of IEEE Visualization Conference Minneapolis, MN, IEEE, Los Alamitos (2005)
- Wilkinson, L.: *The Grammar of Graphics*, 2nd edn. Springer, Heidelberg (2005)
- Zhu, Y., Suo, X., Owen, G.S.: Complexity Analysis for Information Visualization Design and Evaluation. In: Proceedings of the 3rd International Symposium on Visual Computing (ISVC). LNCS, vol. 4841, Springer, Heidelberg (2007)