Data Management Systems

Course Number: 610

Course Term: Summer, 2020

Assignment number: 4

Project title: Dashboard and Story Development using Cognos Analytics

Contact Information:

Melissa Hunfalvay

Email: [melissa@righteye.com](mailto:melissa@righteye.com)

Phone: 240-357-0026

Instructors Names:

Dr. Moretto

Dr. Goldberg

Introduction

The data set chosen for this assignment is Billionaires data (see figure 1). This data set represents a multi-decade database of the super-rich. This dataset was gathered from Forbes World’s Billionaires list from 1996-2014 and Peterson Institute for International Economics have added more information about each person on the list. An example of variables added by the Peterson Institute includes, whether the billionaires were self-made or inherited their wealth.

The original data set, prior to any cleaning contains 2,614 number of cases or rows of data. There are 22 columns (variables) of data. Variable include strings, floats, Boolean and integers. Only minimal and obvious data cleaning was conducted and is described in Assignment 3. After cleaning was complete there remained 2,614 number of cases or rows of data. There are 20 columns (variables) of unique values with an average of $3.53 worth in billions. Therefore, there is a significant amount of unique data to make the dataset robust for exploration.

The dataset provides the following variables on billionaires; age (before cleaning -1 to 98, after cleaning 12 to 98); gender (9.5% female, 90.5% male); the person’s name; their wealth rank; whether they inherited the wealth or not; the region where they live, if they were a founder or were political. Variables also include the wealth type, worth and year the data was reported. Predictive variables related to the citizenship of the individual was included with the largest volume of billionaires from the United States (n = 903) and then Russia (n = 119).

Details on the company were also reported and include company name, date the company was founded and the country code, region, category, sector, type, and year the company was formed. For instance, in the variable category the following sectors are reported; finance sectors (30.6%), non-traded sectors (22.8%), traded sector (21.6%), new sector (12.2%), resource sector (9.4%).

The intent of the analysis for assignment 3 was to find which variables within the dataset predicted billionaire status. The goal of assignment 4 is to use the dashboards to dig deeper into the insights gained from the predictive analysis with new insights revealed via dashboard visualizations and storyboards.

This dataset is an important topic, aside from others wanting to be billionaires because it can be used to explore the disparity of wealth across the world. For instance, the world population, as of this writing, is almost 7.8 billion people (<https://www.worldometers.info/world-population/>) with 2,095 being billionaires (<https://www.forbes.com/billionaires/>). Therefore, according to these figures, 0.00000269 percent of people are billionaires. These 2,095 billionaires collectively have more wealth than 60% of the rest of the worlds population (<https://time.com/5768346/billionaires-wealth/#:~:text=Population%2C%20Report%20Finds-,Billionaires%20Have%20More%20Wealth%20Than%2060,the%20World's%20Population%2C%20Report%20Finds>) underscoring the degree of global inequality. Hence, this dataset can be used to create awareness of such inequalities and spur innovation and social programs to assist others who are less fortunate.

Dashboard Assembly, Presentation and Discussion

The predictive analysis conducted in assignment 3 revealed, in summary, a predictive model (specifically model number 3) to be the most robust with a high predictive score (78%) and a variety of drivers. Model three has the highest predictive value of rank <148 at 6.55. Rank was the sole driver and top rule. The second rule included rank between 148 and 311 as well as country, a combination of two drivers with a predictive value of 4.40%.

Broad similarities were found between the four different models created in assignment 3. First, the top predictive rule was consistently rank. Year and country also contributed to the top five predictive variables for each model. When combining these drivers together they added predictive strength to worth in billions. Inputs that consistently did not contribute to models included gender, age, name, and was political.

Therefore, using dashboards, two additional goals were set to understand the data further. The first was an exploration of countries and the billionaires within each country. The second, was to create an easy way to explore key variables with each individual billionaire to determine individual ‘profiles’ or characteristics. These goals resulted in two very different dashboards. The first is colorful and creative with dynamic updating (see Figure 2), the second is highly informative, with key insights, variables and data points that create a profile of each individual billionaire (see Figure 3).

*Countries and Billionaires Dashboard.* Using this dashboard and building on the predictive driver of countries from assignment 3, the following additional insights were revealed. First, it is clearly visual that the United States had the greatest count of billionaires (see figure 2, word scramble of number of billionaires per country). Second, the United States billionaires were worth more, collectively, than any other country (see figure 2, worth in billions by country).

A third additional insight from this dashboard was worth in billions by wealth type. When filtering on the United States only (see figure 4) billionaires were most prevalent when coming from a founder, non-finance background than other backgrounds. Inherited wealth was the second wealth type with the largest worth in billions. This is consistent with the Forbes article about keeping wealth all in the family (<https://www.forbes.com/sites/learnvest...>).

A fourth insight from this dashboard revealed wealth type by industry sector (figure 5) revealing that media moguls, followed by real estate, then retail sectors within the United States were most frequent when associated with wealth type. This kind of insight can a) help further explain the country driver within the model and, b) can also create more questions, filters and insights that could strengthen the model. For instance, when reviewing wealth type there are two medium size categories that seem highly related (technology and software) and could be merged to form one variable which may influence the models’ stability and predictability.

*Billionaire Individuals Profile Dashboard.* This dashboard (figure 3) was created to explore individuals within the dataset against key variables that may a) influence their billionaire status, b) provide a place where personal characteristics related to individuals can be easily visualized creating profiles of each person, c) a way to look across individuals with a play bar option to determine if certain trends are seen from a new visualization and storyboard perspective. This information may be used to further explore model development or simply to provide insights on a more granular level related to individuals rather than groups.

This dashboard displays sum of money across the dataset, maximum value, minimum value, and number of distinct records. For instance, when filtering on Bill Gates there are 3 distinct records. This means there is data on Bill gates across three years (see figure 6). Donald Trump has 2 records (figure 7), however, when using the play feature, it is interesting to watch that most people only have one record (e.g. Albert Ueltschi). This is an important additional insight as it reveals influence of the predictive value of year. Specifically, if a person did not have a billionaire status (or did not have data within a specific year) then the data set and its related outcomes may be skewed. In a similar vein, if one person had multiple data sets (e.g. Bill Gates) then the demographic and epidemiological variables associated with Bill Gates would be ‘counted’ three times if no filter was added to the variable of year. This would skew the data set towards the United States (as Bill Gates is a United States citizen) as having a higher predictive value simply because one person has multiple entries. Such insights can lead to further predictive model refinement by specifically viewing data models by year or removing people with multiple data entries.

Additional insights are revealed in the sum of money component of this dashboard (see figure 3, 6, 7). This widget adds the total sum of money for the individual billionaire. A further exploration of this data, and perhaps a more robust model, could be explored by bifurcating the data based on a natural cut-off within a distribution plot. For instance, if there is a bi- or multi-model distribution based on total sum of money, it may be interesting to create clusters or groups around this metric to see if a more robust or stable model is revealed.

In summary, dashboards provide additional insights via visualizations to allow for a) a better understanding of this dataset, b) thought provoking ways to refine, create and explore predictive models related to the data set.

Story Assembly Discussion and Display

Storyboards were developed for both dashboards. Features within the storyboards include a) timing of the segments within the dashboard; b) sequences of different components of the dashboards; c) features that were permanently displayed (e.g. name) and others that were introduced and removed, still others that built on one another, showing more information over the story. The design of each storyboard was carefully considered to walk thru the data emphasizing a certain focused flow for the viewer in order to tell a compelling story.

*Countries and Billionaires Dashboard, Story 1:* The key features of this storyboard reveal information about the location of billionaires. Where are these people located? How many are in each country? What made them wealthy? What industries are billionaires doing business in? The key features this storyboard creates includes building upon the general characteristics associated with these billionaires (see figure 2). This dashboard has a lot of color and includes all the billionaires across the different variables. Therefore, it can be difficult to dig deeper without a step-by-step serial or storyboard view of each variable. The story provides this context, giving the viewer additional insights.

In summary, conclusions can be drawn by focusing on worth in billions by wealth type (figure 4) where the data reveals how this one variable influences all billionaires in the data set. Conclusions can also be drawn from wealth rank by industry/sector (figure 5), with real estate being the largest contributor. Providing the data in this manner, allows a storyboard with a lot of color and information to be viewed differently to provide additional insights.

Future development of this storyboard could be to filter the data by country and have all the follow-on visuals relate to one country at a time. This may also, provide insights to either group certain countries together, or look at specific features within a country. These additional insights may help to explain how billionaires became so wealthy within a country’s economic infrastructure.

*Billionaire Individuals Profile Dashboard, Story 2:* The key features of this story board involve a ‘walk-thru’ profile of individuals within the data set (see figures 3, 6, 7). This begins with the individuals name, then the number of unique data sets. These two features tell the viewer a) who the person is; b) how much information we have on them. Then the story moves towards describing the individual starting with their gender, then what category of business they are associated with. Finally, the storyboard reveals the individuals’ total worth, their minimum and maximum worth values.

This storyboard presents who the person is, then a little about them as an individual and finally, their financial status. By linking this information together in a storyboard fashion, the individual is ‘presented’ to the viewer. He or she is humanized and not simply a number on a spreadsheet. These storyboard insights can assist in the viewer in understanding the individuals and ‘getting to know them’ in a sequence, rather than in dry, interdependent pieces of data.

In conclusion, let me tell you the story of Bill Gates (see figure 6), he has three unique data sets, in other words, he had billionaire status across three different years. He is a male and his business is in a new sector, so he is likely an innovator. His minimum wealth was 18.5 billion and his maximum wealth was 76 billion, his total wealth was 153.2 billion. Bill Gates is indeed a very wealthy man!

Impact on your Organization

Providing data in a visual fashion via a dashboard would help my organization tremendously. First, RightEye has a lot of data and variables. Without visual representation and storyboards it is difficult to discern trends, and insights. Second, those within the organization who are not data experts do not know how to make sense of data without ‘seeing’ it. By seeing it, I do not mean in the form of spreadsheets. Instead, clear visuals that everyone can relate to and understand are important for company communication.

Providing data in the form of a storyboard can be very powerful. For instance, the second dashboard that profiles each individual was created with the RightEye Impact Awards in mind (<https://righteye.com/impactaward/#:~:text=The%20RightEye%20Impact%20Award%20is,lives%20through%20functional%20vision%20healthcare.>) The purpose of the impact awards was to ‘profile’ specific patients who benefited from the RightEye product. Using the individual storyboard profile, I could highlight these individuals giving them a name, face, changes in symptomology, changes in data results, vision therapy sessions, pre and post individual changes. I could further humanize their stories by adding videos and testimonials. This is important to my organization from every aspect. First, it tugs at everyone’s heart strings, and boosts moral of RightEye employees, customers, patients, and potential users can learn how RightEye can improve patients’ lives. Second, it helps with marketing and public relations. Third, it is consistent with the organizations goal of being mission based. Fourth, such stories can we used in presentations for investors, interested future clients and organizations. Everyone wants to be associated with an organization that is doing good work and making a difference. Using storyboard profiles can provide a beautiful way to visualize the companies’ mission.

A possible disadvantage to such visualizations and storyboards in my organizational context is that this data and these patient stories are very personal. We need to obtain the correct privacy consents and display the data and stories with the dignity these individuals deserve, to show the courage and struggles they have had to endure.

A second con to the use of such visualizations in my organization is how to make this kind of story scalable and easy for our customers to create. There is no question in my mind that the information is valuable. However, with one (very expensive) license, how do I allow these features to be available to all customers? This is a very practical concern for a small business.

In summary, the purpose of this project was to create dashboards and storyboards from data examined in assignment 3. Two goals were set to understand the data further. The first was an exploration of countries and the billionaires within each country. The second, was to create an easy way to explore key variables with each individual billionaire to determine individual ‘profiles’ or characteristics. These goals resulted in two very different dashboards that provided additional insights into the dataset to a) understand the data further, b) refine the predictive model, c) tell a more understandable story allowing the entire company to understand the insights from the dataset.

References

Bates, J. (2020). *Time Magazine.* Retrieved from <https://time.com/5768346/billionaires-wealth/#:~:text=Population%2C%20Report%20Finds-,Billionaires%20Have%20More%20Wealth%20Than%2060,the%20World's%20Population%2C%20Report%20Finds>

Barclay, M. (2019). *RightEye.* Retrieved from <https://righteye.com/impactaward/#:~:text=The%20RightEye%20Impact%20Award%20is,lives%20through%20functional%20vision%20healthcare.>

Bianchi, J. (2015). What you can learn from the 1% about keeping family wealth going. *Forbes Magazine.* Retrieved from: <https://www.forbes.com/sites/learnvest/2015/07/30/what-you-can-learn-from-the-1-about-keeping-family-wealth-going/#2d1f0bf794ed>

Dolan, K.A. (2020). *Forbes World’s Billionaire List.* Retrieved from <https://www.forbes.com/billionaires/>

Regent, D.(2020). *Worldometer.* Retrieved from <https://www.worldometers.info/world-population/>

Appendix

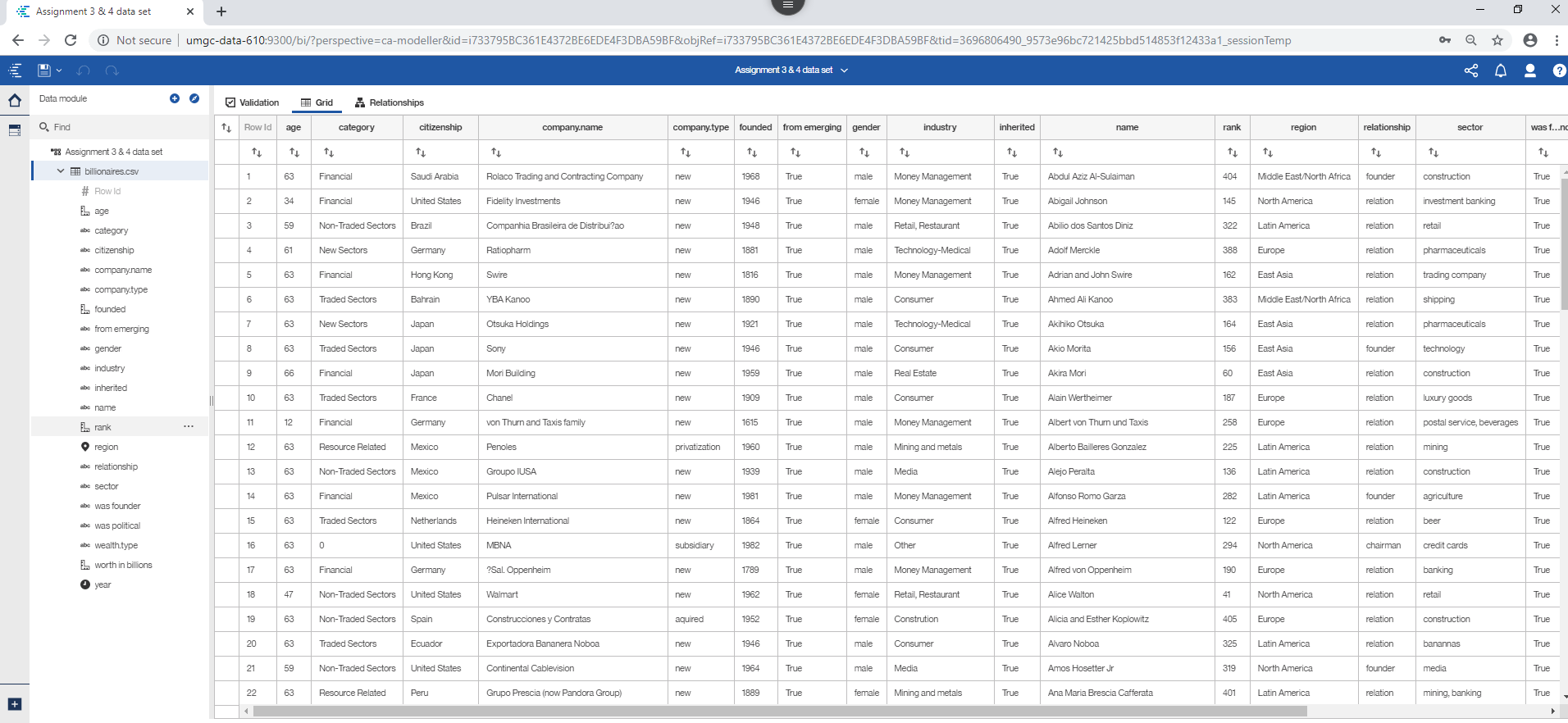


Figure 1: Partial screenshot of the data set

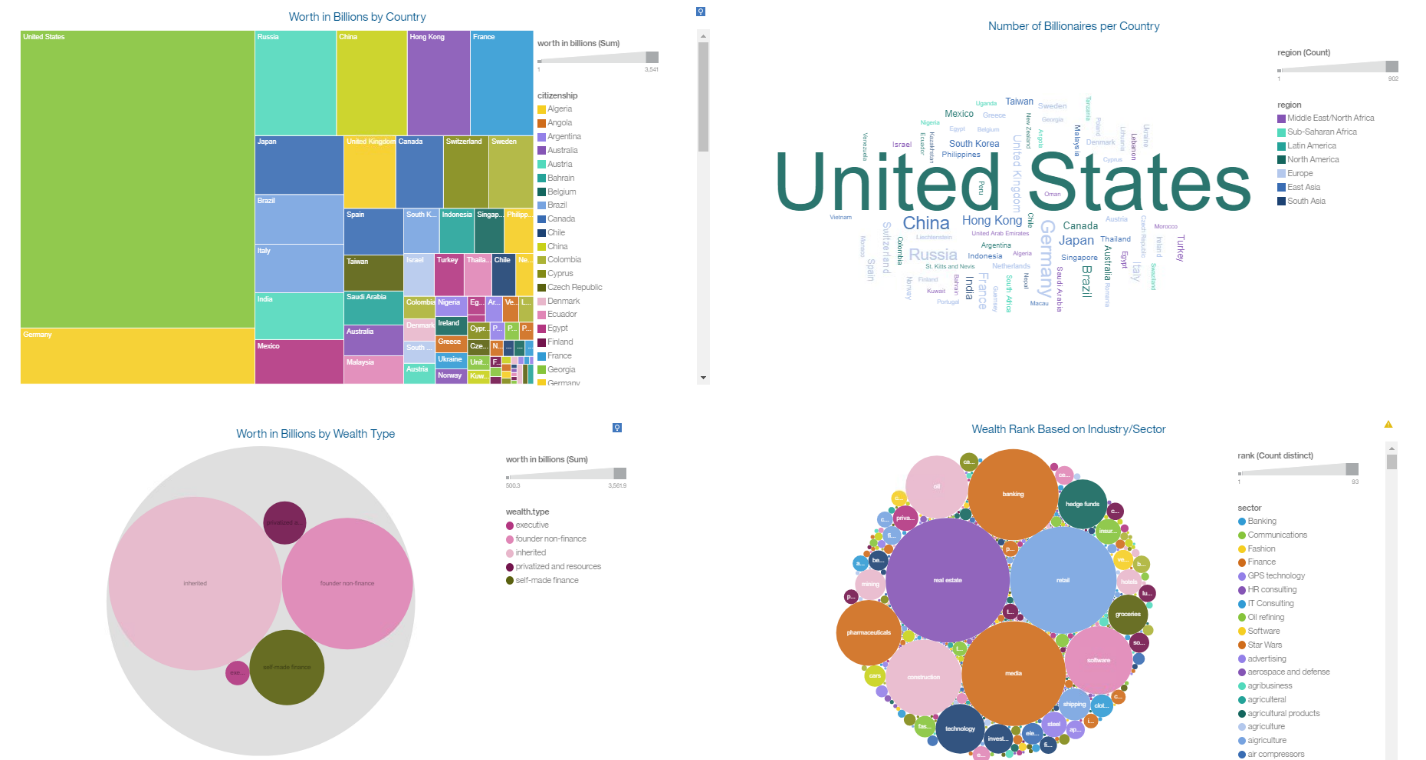


Figure 2: Creative global dashboard based on billionaires within countries. No filters.

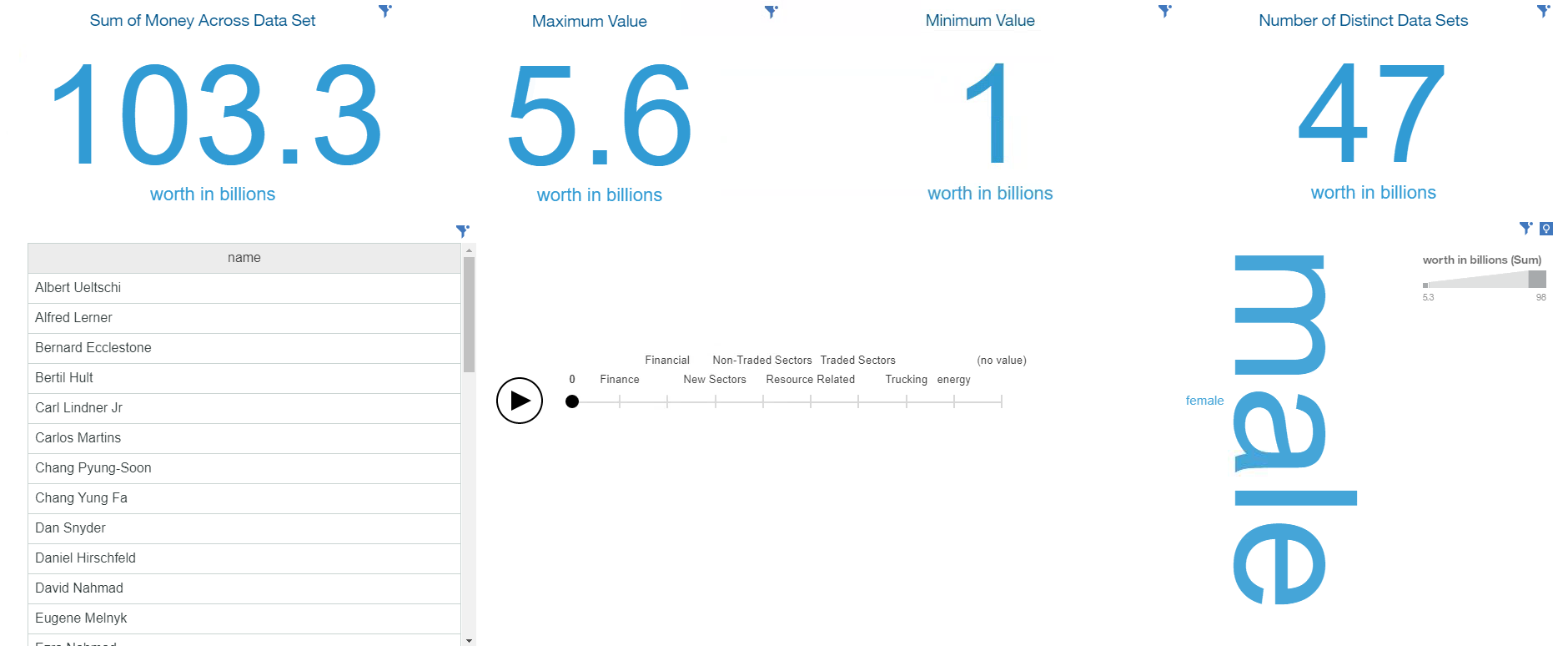


Figure 3: Dashboard profiling individual billionaires

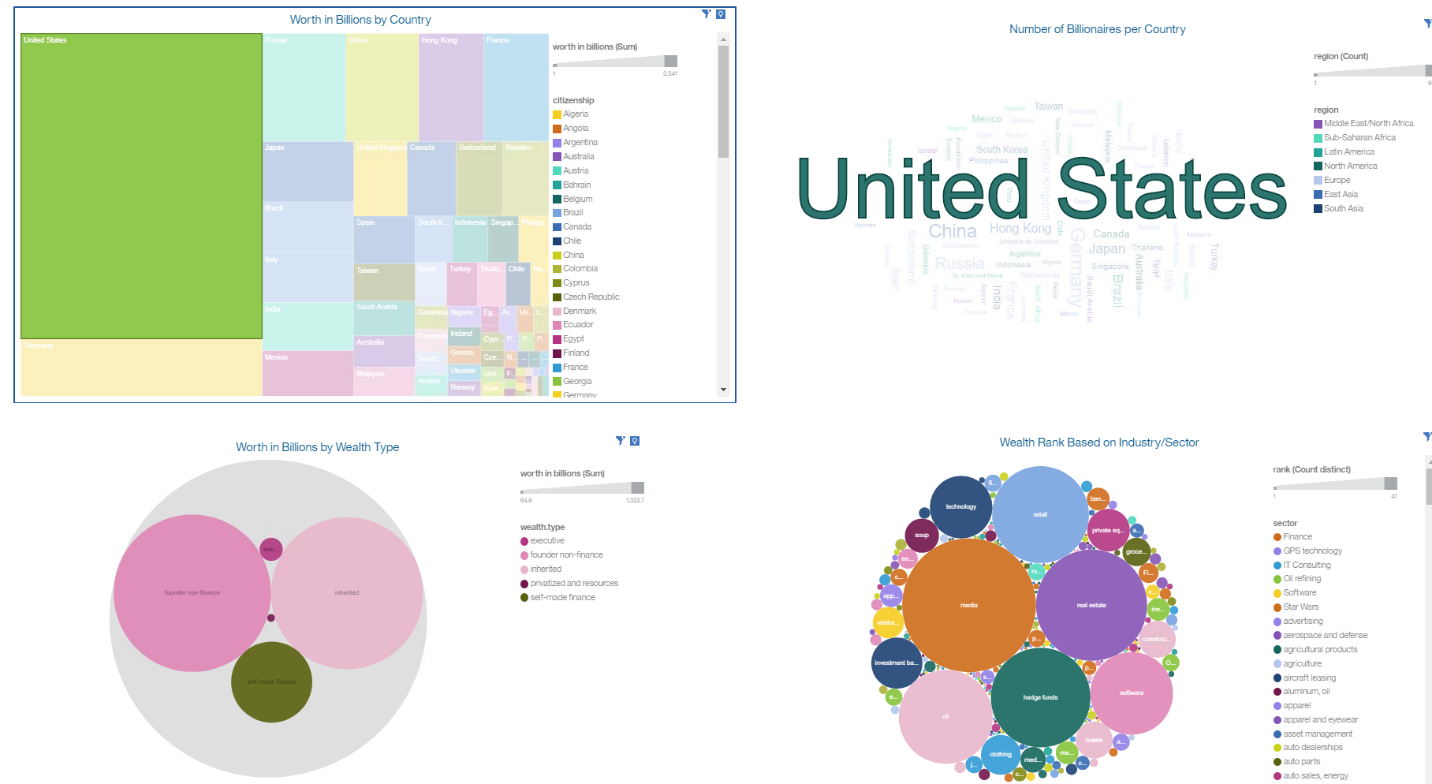


Figure 4: Filtering based on the United States worth in billions by wealth type

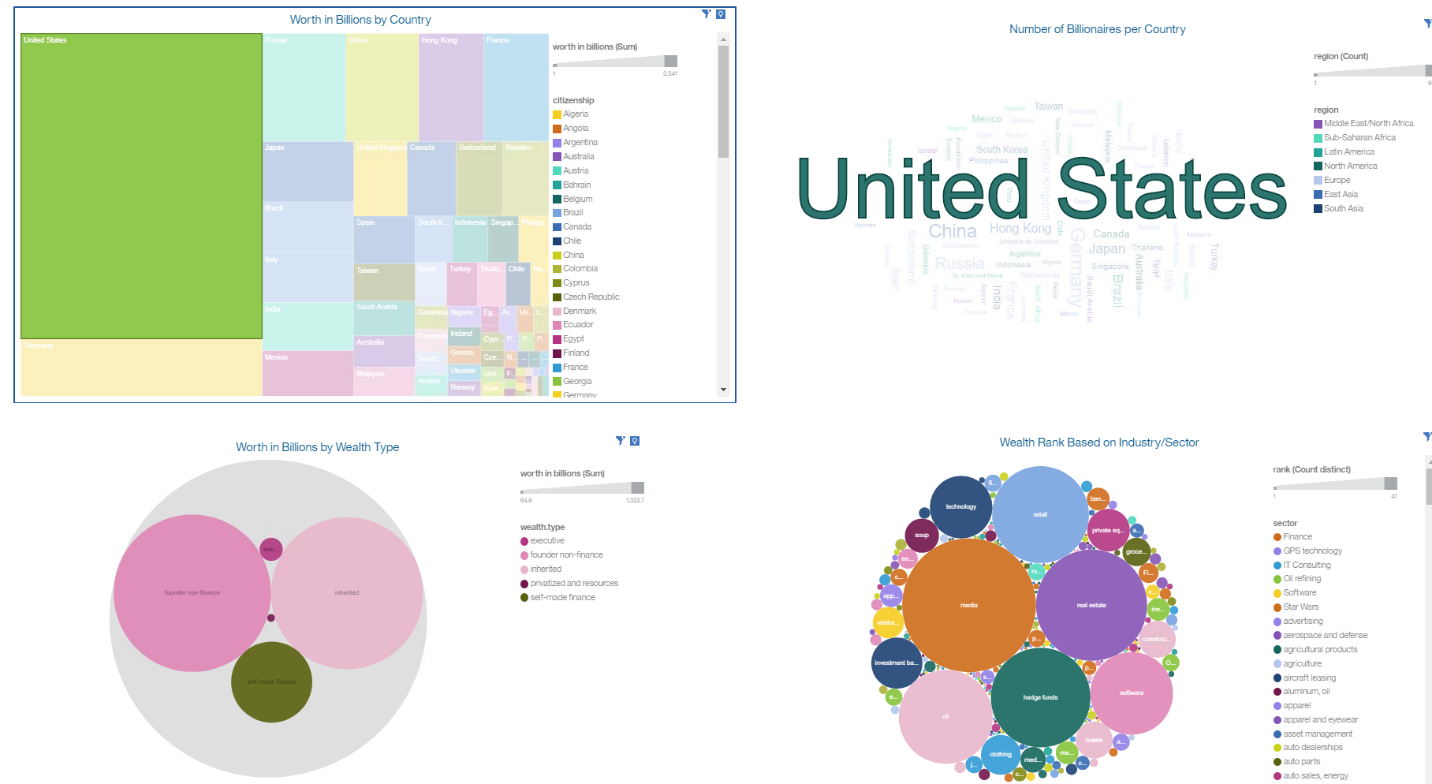


Figure 5: Filtering based on the United States wealth rank based on industry sector

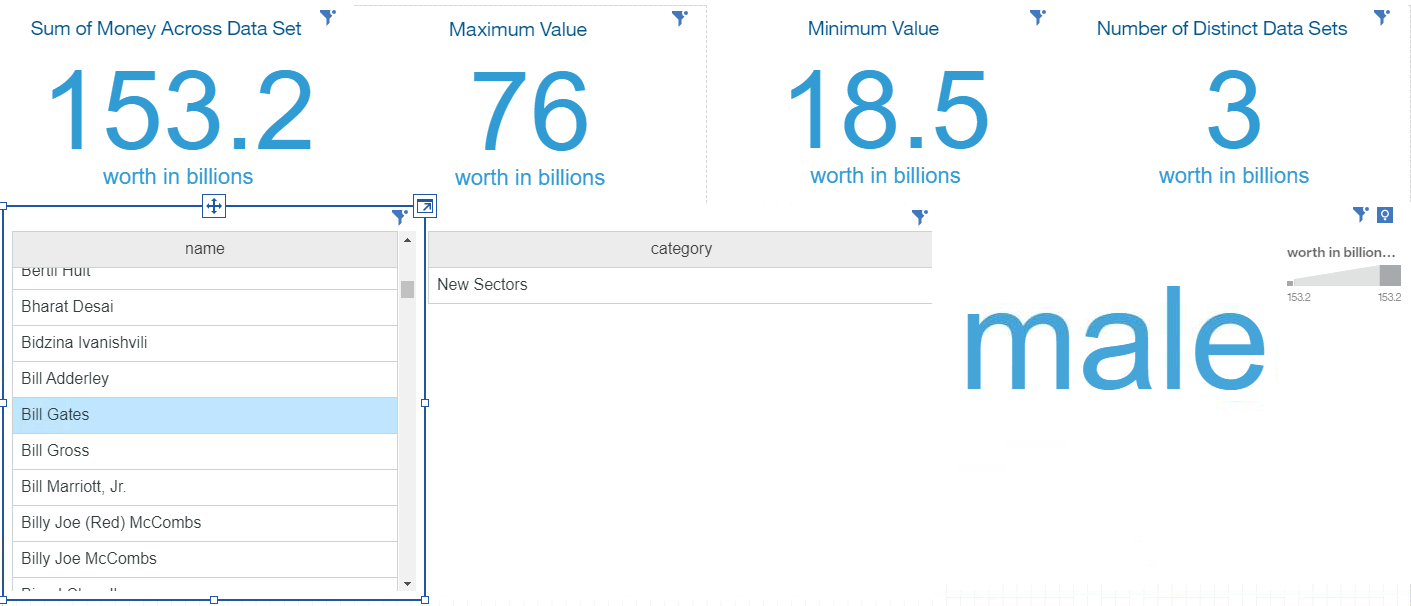


Figure 6: Bill gates billionaire profile

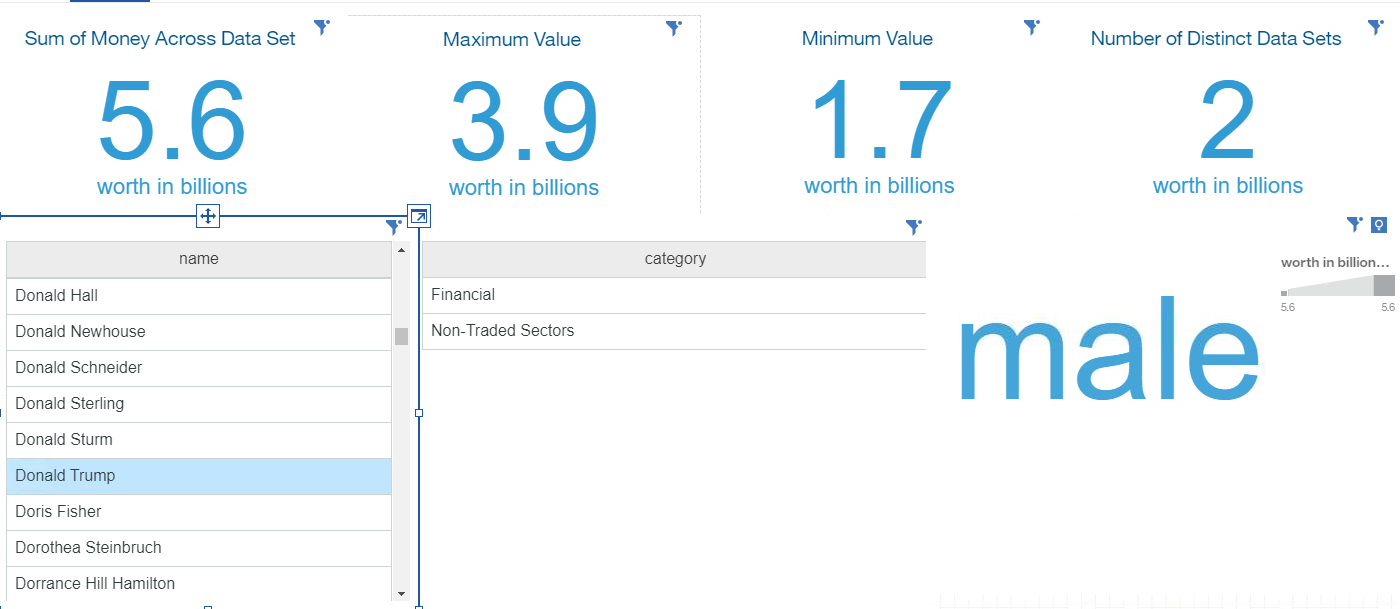


Figure 7: Donald Trump billionaire profile