Assignment 4: Display Development Using Watson Analytics

Daanish Ahmed
DATA 610 9080
Fall 2016
dsahmed2334@yahoo.com
Professor Vahe Heboyan
November 27, 2016

Introduction

Implementing data analytics into the decision-making process requires more than just analyzing data. Data analysts must be able to effectively communicate their findings to an audience in order for those findings to bring action. According to Davenport (2013), an effective presentation involves telling a story with the data that audiences can understand. To make this process easier, IBM's Watson Analytics features a display tool which allows users to present their findings in an organized and visually-appealing way. My goal in this assignment is to communicate the results of my analysis through the use of Watson Analytics' display features. Using my findings from two previous assignments, I hope to be able to tell a story with the data that would have otherwise been difficult to do without the use of displays.

For this assignment, I used two datasets obtained from IBM's list of sample datasets—one focuses on employee attrition while the other revolves around sales profits and costs (Stacker, 2015). The dataset on employee attrition, used in my second assignment, is fairly robust with about 1470 cases and 35 possible input values (see Table 1 in Appendix A). Some of its noteworthy inputs include education field, job role, years working, age, and distance from home to work. The variables that can be used to develop predictive models include employee attrition, job role, and the number of companies an employee worked for. The dataset on sales, used for the third assignment, is very robust since it features 24,743 cases and 14 possible inputs (see Table 2 in Appendix A). A few of the major input values include revenue, product type, quantity, order method type, product cost, and retailer country. Predictive models can be developed from gross profit, quantity, and revenue. Both of these datasets were cleansed of outliers, errors, and missing values in the previous assignments, and my displays will use the refined versions of these datasets. In this paper, I will create two dashboard displays based on the results obtained from previous assignments—the dataset on employee attrition will have attrition as its output variable while the sales dataset will focus on gross profit. Afterwards, I will implement data from social media and incorporate a storybook display to enhance my analysis. Finally, I will look at how this approach of communicating the results can be useful for solving problems in my organization.

Employee Attrition Dashboard

In this section of the paper, I will describe the display that I created from the employee attrition dataset. I had previously studied the variables influencing attrition and found that they included job role and age. This display is designed to describe the relationship between these two inputs and show their specific impact towards attrition (see Figure 1 in Appendix A). The first image of this display shows a pie chart featuring a breakdown of attrition cases among employees based on their job role. According to this graph, the roles with the highest number of cases are laboratory technicians, sales executives, and research scientists. The second visualization is a stacked bar graph showing the percentages of attrition for each job role, as opposed to the total number of overall cases as seen in the first image. Here, the roles with the highest attrition rates are sales representatives, laboratory technicians, and human resources. At first, these results may appear different from those of the first image. For instance, sales representatives have by far the highest rates of attrition, even though they make up less than 14% of all attrition cases. But by using an additional bar graph, I found that there are far fewer sales representatives than there are sales executives or laboratory technicians (see Figure 2 in Appendix A)—thus explaining why those two roles have a higher number of attrition cases. These findings are useful for my organization because they can help the company to pinpoint which employees need the most attention. These visualizations reveal not only which groups are most likely to churn, but also the probability of attrition within each role.

The third image shows an area curve of the employee attrition rate according to age. The results indicate that attrition is highest for younger and less-experienced employees. According to Roberts (2015), workers are indeed more likely to leave their jobs within their first few years of working. The fourth image of this display is a packed bubble visualization showing the average age of workers with attrition for each job role. The diagram indicates that employees working as sales representatives, human resources, and laboratory technicians have the youngest average age at around 28 to 31 years old. These three job roles also have the highest attrition rates, making these results consistent with my findings. Job roles with the highest average employee age—such as managers and research directors—likewise have the lowest attrition rates. Ultimately, these results suggest that job roles held by more experienced workers generally result in lower attrition rates. Thus, my organization is more likely to see lower attrition levels if it can provide to the needs of less-experienced workers holding certain job roles. In the end, this dashboard helps to improve the depth of my previous analysis on employee attrition. My earlier models focused on the individual factors influencing attrition, but this display shows how strongly this combination of variables can affect the output.

Gross Profit Dashboard

After looking at the employee attrition display, I will now discuss the results of creating a dashboard based on the sales dataset. In my previous assignment, I used this dataset to develop a predictive model for gross profit, and I found that product type and retailer country have a significant impact on profit. For this display, I intend to tell a more specific story regarding the profitability of different product types across certain countries (see Figure 3 in Appendix A). By looking at the visualizations in this display, many useful insights can be gained. The first image shows a word cloud listing the retailer countries based on profits generated from sales in each country. Some of the highest-grossing countries include the United States, Japan, China, Canada, and France. The second image offers further insight by featuring a series of pie charts showing the preferred order method types within the ten highest-grossing nations. This visualization shows firstly that the internet is the most popular method for ordering products in all ten of these countries. But still, some of the preferred order method types can vary drastically in each country. For instance, customers in China and Korea prefer ordering through the web more than using all other methods combined. Meanwhile, countries like Canada and Germany have at least a quarter of consumers ordering through email, while virtually none use this method in China. By showing the preferred order methods in each country, this image provides vital information to my organization because it can allow the company’s marketing team to determine the best approach for promoting products in each country.

The third image in this display is a packed bubble visualization that shows different product types according to their respective gross profits. Some of the most successful product types include eyewear, watches, tents, packs, woods, and sleeping bags. The fourth image is a treemap that shows the performance of each product type in different countries. It lists the ten highest-grossing countries and the profit generated in each country by the top eight product types. Interestingly, the ranking of the product types is relatively similar across most of these nations. For instance, eyewear is the most profitable product in all ten countries. However, there is some degree of variability regarding the ranking of these product types in some nations. For example, cooking gear ranks higher in sales within the U.S. than in any other country. Also, consumers in Korea have a higher preference for tents than those in other countries, while Chinese customers are less likely than others to purchase sleeping bags. These insights can greatly benefit my organization by revealing which product types will be the most successful in each country. Although most of these countries' consumers have similar preferences, it is important to know the differences so that my company can target the right audiences. Overall, this dashboard improves my previous analysis by showing a more specific look at some of the factors influencing gross profit. My earlier model yielded a general list of decision rules for my organization to follow, but this display offers a clear sense of direction for how my company can effectively market its products internationally.

Social Media Aspect

After creating these two displays, I chose to enhance my analysis through the use of social media. I imported a dataset from Watson Analytics consisting of Twitter tweets based on the top product types from the sales dataset (see Table 3 in Appendix A). Some of the hashtags referred to products such as eyewear, packs, and sleeping bags. From here, I created a dashboard of my analysis (see Figure 4 in Appendix A) to compare my findings with those listed in the gross profit dashboard. The first image in this display shows a world map indicating the number of tweets per country using the specified hashtags. The results show that the majority of tweets come from the United Kingdom, United States, and Italy. The second image on the left is an area curve showing the number of tweets per day according to their product hashtags. This graph shows that the grand majority of tweets include hashtags about watches. However, the gross profit dashboard suggested that watches are only the second highest-grossing product type—below eyewear. The graph also shows that the number of tweets peaked on November 23 and 25. These peaks coincide with the influx of purchases that occur on the day before Thanksgiving as well as on Black Friday. Therefore, these results are helpful because they provide some insight on when consumers are most likely to make purchases.

The image on the right side of the display is a stacked bar graph that shows the most common product-related tweets for each author's country of origin. Based on this graph, watches still make up a large majority of hashtags in most nations. However, eyewear also receives frequent tweets—especially in countries like Austria and Japan. Likewise, users in Germany show a strong preference for packs, while Dutch and Spanish users frequently tweet about woods. As a result, these tweets may provide some insight into the minds of consumers living in many of these countries. However, I believe that my organization should take great caution when using this approach because it is not guaranteed that tweets containing these hashtags are referring to purchases or customer interests. Additionally, the tweets sampled in some countries may not represent those countries' populations due to the fact that certain nations have very few tweets when compared to others. Altogether, these issues may also explain why the results in this visualization differ from those in the gross profit dashboard.

Storybook Display

The final part of my exploration consists of implementing a storybook display into this assignment. Here, I created a new storybook using the three displays that I designed so far (see Figure 5 in Appendix A). Using comments and annotations, I provided descriptive hints to viewers in order to help them understand my analysis (see Figures 6, 7, and 8 in Appendix A). The goal of my storybook was to show business owners a few ways in which they can reduce employee attrition and increase profits in their company. One insight from using this format is that it makes the individual parts of my analysis come together as a whole. This storybook allows users to interact with the data and easily identify key takeaways for improving several aspects of their businesses. Users can even replace my data with one of their own datasets. However, I would use caution before taking this approach. I also explored the ability to substitute my dataset into an existing storybook template by using the sales effectiveness template from IBM's Analytics Exchange ("Analytics Exchange," 2016). I replaced columns such as product group and territory name with product type and retailer country, respectively. However, the resulting displays provided very little useful information and were often filled with errors (see Figure 9 in Appendix A). This happened because many of the replacement columns did not match well with the original columns. Therefore, I believe that users should only import templates if their dataset is very similar to the original dataset used for that template.

Applications

As I mentioned throughout this paper, there are many ways in which my organization can benefit from implementing displays. One possible application is to use analytics for identifying which employees are most likely to churn. By using a display, I was able to show that younger workers with certain job roles were more likely to leave their jobs. This methodology allowed me to communicate this information to my audience so that they can easily identify important sources of attrition. Similarly, this approach can easily be applied towards maximizing profits across my organization's international markets. My displays revealed the preferences of each country with regards to products and ordering methods, as well as the times when consumers are most likely to make purchases. The benefit of using this approach is that it allows others to easily understand some of the causes for these problems. As a result, my company can quickly develop specific solutions without having to spend too much time researching the problems.

However, this display-based approach is not without its issues. One of the shortcomings of using my method is that the displays have a relatively narrow focus. In my previous assignment, I found that using decision trees yielded numerous variables influencing the target; but the displays I created here focused mainly on one or two inputs each. Although displays can involve many variables, they tend to work most effectively when having a specific focus. Multiple displays might have to be created in order to map out solutions for each important variable, which can be a time-consuming process. In addition, the social media features provided by Watson Analytics may not always yield the most accurate results if there is a shortage in the number of relevant tweets about the subject. But in the end, I believe that any shortcomings of using my approach can be minimized if my organization uses great care when implementing these features. If my company were to adopt this display-based approach, then it is very likely that the results will lead to stronger decision-making.

References

Analytics Exchange - Sales effectiveness. (2016, October 18). Retrieved November 25, 2016, from https://watson.analytics.ibmcloud.com/marketplace/details/5ee6e61b152b698900718439ac65244e

Davenport, T. (2013, January 31). Telling a story with data. *Deloitte University Press*. Retrieved November 20, 2016, from http://dupress.com/articles/telling-a-story-with-data/

Roberts, P. (2015, January). The CFO and CHRO Guide to Employee Attrition. *Workforce Solutions Review,* 6(1), 8-10. Retrieved October 28, 2016, from UMUC Library.

Stacker, M., IV. (2015, April 2). Guide to Sample Data Sets. Retrieved October 25, 2016, from https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets//

Appendix A

EducationField	EnvironmentS...	JobSatisfaction	Attrition	EducationLevel	JobRole	MaritalStatus	OverTime	Gender	Age	TotalWorkingY...	YearsAtComp...	YearsInCurrent...	TimeWithCurr...	NumCompani...	JobLe
Medical	1	3	No	4	Research Dir...	Married	No	Female	47	26	20	15	5 to < 7	3	
Medical	4	3	No	1	Sales Execut...	Married	Yes	Female	24	5	5	4	2 to < 5	1	
Marketing	4	4	No	5	Sales Execut...	Married	No	Male	32	7	4	3	2 to < 5	4	
Life Sciences	4	2	No	3	Laboratory T...	Married	No	Female	34	7	5	4	2 to < 5	5	
Medical	2	3	No	5	Research Sci...	Married	Yes	Male	41	7	4	2	2 to < 5	2	
Other	3	3	No	4	Laboratory T...	Divorced	No	Male	40	11	8	7	7 to < 9	3	
Life Sciences	1	4	No	2	Sales Execut...	Divorced	No	Male	31	13	13	7	9 to < 12	1	
Marketing	1	4	Yes	3	Sales Execut...	Divorced	No	Male	46	28	7	7	2 to < 5	4	
Life Sciences	1	1	Yes	3	Laboratory T...	Single	Yes	Female	39	11	1	0	less than 2	2	
Life Sciences	3	2	Yes	5	Manufacturin...	Single	No	Female	31	10	10	8	7 to < 9	1	
Medical	2	1	No	3	Healthcare R...	Divorced	Yes	Male	45	24	7	7	7 to < 9	2	
Medical	4	2	No	2	Human Reso...	Single	No	Female	31	8	3	2	2 to < 5	9	
Life Sciences	2	4	Yes	3	Laboratory T...	Married	Yes	Male	31	7	2	2	2 to < 5	6	
Technical De...	2	4	No	3	Manufacturin...	Married	Yes	Male	45	10	3	1	2 to < 5	4	
Marketing	3	1	No	3	Sales Execut...	Divorced	No	Male	48	15	2	2	2 to < 5	4	
Technical De...	1	3	Yes	4	Human Reso...	Married	No	Female	34	2	2	2	2 to < 5	1	
Medical	4	2	No	1	Research Dir...	Divorced	No	Male	40	16	9	8	7 to < 9	2	
Medical	4	1	No	3	Sales Execut...	Single	No	Male	28	6	5	4	2 to < 5	0	
Life Sciences	3	3	No	3	Laboratory T...	Single	No	Male	44	7	5	2	2 to < 5	3	
Medical	4	2	No	3	Research Dir...	Single	No	Male	53	34	9	8	7 to < 9	4	
Technical De...	1	3	No	4	Healthcare R...	Married	No	Male	49	20	3	2	2 to < 5	2	
Medical	3	3	No	3	Research Sci...	Divorced	No	Male	40	8	3	1	2 to < 5	2	
Life Sciences	4	2	No	3	Research Sci...	Single	No	Male	44	6	5	2	2 to < 5	1	
Medical	4	1	No	3	Sales Execut...	Married	No	Male	33	5	4	3	2 to < 5	0	
Other	4	1	No	3	Sales Execut...	Single	No	Male	34	15	13	9	12 and above	3	
Life Sciences	4	3	No	1	Sales Execut...	Married	No	Male	30	4	2	1	2 to < 5	8	
Medical	1	4	No	2	Laboratory T...	Single	No	Female	42	12	12	9	7 to < 9	1	
Marketing	1	3	No	5	Sales Execut...	Married	No	Female	44	11	1	0	less than 2	7	
Technical De...	3	3	No	3	Research Sci...	Divorced	No	Male	30	1	1	0	less than 2	1	
Life Sciences	2	3	No	2	Research Sci...	Married	No	Male	57	13	12	9	7 to < 9	0	
Life Sciences	3	2	No	4	Healthcare R...	Divorced	No	Male	49	29	8	7	7 to < 9	3	
Medical	2	1	No	3	Research Dir...	Divorced	No	Male	34	16	14	8	9 to < 12	7	
Technical De...	1	3	Yes	3	Sales Repres...	Married	No	Male	28	5	3	2	2 to < 5	3	

Table 1. Employee Attrition Dataset.

Product line	Product type	Product	Order method ...	Retailer country	Year	Quantity	Gross profit	Planned revenue	Revenue	Unit price	Unit sale price	Pr
Camping Eq...	Lanterns	Firefly 2	Mail	Canada	2005	1171.0	$11,779.27	$32,050.27	$31,299.84	$27.37	$26.73	
Camping Eq...	Lanterns	Firefly 2	Mail	Japan	2005	22.0	$276.98	$602.14	$602.14	$27.37	$27.37	
Camping Eq...	Lanterns	Firefly 2	Mail	Netherlands	2005	902.0	$9,252.90	$24,687.74	$24,213.64	$27.37	$27.00	
Camping Eq...	Lanterns	Firefly 2	Mail	Germany	2005	368.0	$4,098.76	$10,072.16	$9,951.71	$27.37	$27.19	
Camping Eq...	Lanterns	Firefly 2	E-mail	Canada	2005	3343.0	$33,088.46	$91,497.91	$88,774.69	$27.37	$26.62	
Camping Eq...	Lanterns	Firefly 2	E-mail	Japan	2005	3477.0	$35,494.07	$95,165.49	$93,298.79	$27.37	$26.92	
Camping Eq...	Lanterns	Firefly 2	E-mail	Sweden	2005	23.0	$289.57	$629.51	$629.51	$27.37	$27.37	
Camping Eq...	Lanterns	Firefly 2	E-mail	Germany	2005	3703.0	$37,357.39	$101,351.11	$98,916.30	$27.37	$26.82	
Camping Eq...	Lanterns	Firefly 2	E-mail	Italy	2005	773.0	$7,845.95	$21,157.01	$20,731.86	$27.37	$26.82	
Camping Eq...	Lanterns	Firefly 2	Fax	France	2005	590.0	$5,988.50	$16,148.30	$15,823.80	$27.37	$26.82	
Camping Eq...	Lanterns	Firefly 4	Telephone	United States	2005	2781.0	$28,713.14	$81,872.64	$76,002.74	$29.44	$27.73	
Camping Eq...	Lanterns	Firefly 4	Telephone	Japan	2005	213.0	$2,311.05	$6,270.72	$6,145.05	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Telephone	Korea	2005	603.0	$5,894.85	$17,752.32	$15,377.25	$29.44	$26.20	
Camping Eq...	Lanterns	Firefly 4	Telephone	China	2005	641.0	$6,954.85	$18,871.04	$18,492.85	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Telephone	Singapore	2005	482.0	$5,229.70	$14,190.08	$13,905.70	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Telephone	Australia	2005	805.0	$8,435.05	$23,699.20	$22,291.45	$29.44	$27.97	
Camping Eq...	Lanterns	Firefly 4	Telephone	Netherlands	2005	831.0	$8,581.15	$24,464.64	$22,617.55	$29.44	$27.53	
Camping Eq...	Lanterns	Firefly 4	Telephone	France	2005	1138.0	$11,700.21	$33,502.72	$30,690.21	$29.44	$27.42	
Camping Eq...	Lanterns	Firefly 4	Telephone	United Kingdom	2005	106.0	$1,150.10	$3,120.64	$3,058.10	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Telephone	Austria	2005	631.0	$6,846.35	$18,576.64	$18,204.35	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Sales visit	United States	2005	513.0	$5,566.05	$15,102.72	$14,800.05	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Sales visit	Mexico	2005	2117.0	$22,323.45	$62,324.48	$59,061.45	$29.44	$28.19	
Camping Eq...	Lanterns	Firefly 4	Sales visit	Korea	2005	330.0	$3,580.50	$9,715.20	$9,520.50	$29.44	$28.85	
Camping Eq...	Lanterns	Firefly 4	Sales visit	Australia	2005	563.0	$6,108.55	$16,574.72	$16,242.55	$29.44	$28.85	

Table 2. Retail, Sales, Marketing Profit-Cost Dataset.

Year (Posted ti...	Month (Poste...	Day (Posted ti...	Hour (Posted t...	Author country 57 of 58	Author state 105 of 106	Author city 250 of 251	Sentiment 4 of 5	#eyewear	#packs	#sleepingbags	#tents	#v
2016	November	24	15	india	nct	new delhi	neutral	0	0	0	0	1
2016	November	24	13	united states	ohio	columbus	neutral	1	0	0	0	0
2016	November	24	1	united states	ohio	columbus	positive	1	0	0	0	0
2016	November	19	2	trinidad and t...	city of port of ...	port-of-spain	ambivalent	0	0	0	0	1
2016	November	22	6	qatar	ad dawḩah	doha	positive	0	0	0	0	1
2016	November	22	16	qatar	ad dawḩah	doha	neutral	0	0	0	0	1
2016	November	23	5	qatar	ad dawḩah	doha	neutral	0	0	0	0	1
2016	November	18	0	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	18	15	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	18	20	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	19	8	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	19	12	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	19	18	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	20	1	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	20	22	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	20	23	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	21	0	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	21	0	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	21	17	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	19	15	united kingdom	england	london	positive	0	0	0	0	1
2016	November	20	2	united kingdom	england	london	positive	0	0	0	0	1
2016	November	18	3	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	18	4	united kingdom	england	london	neutral	0	0	0	0	1
2016	November	18	12	united kingdom	england	london	neutral	0	0	0	0	1

Table 3. Dataset of Twitter Hashtags Containing Top Products.

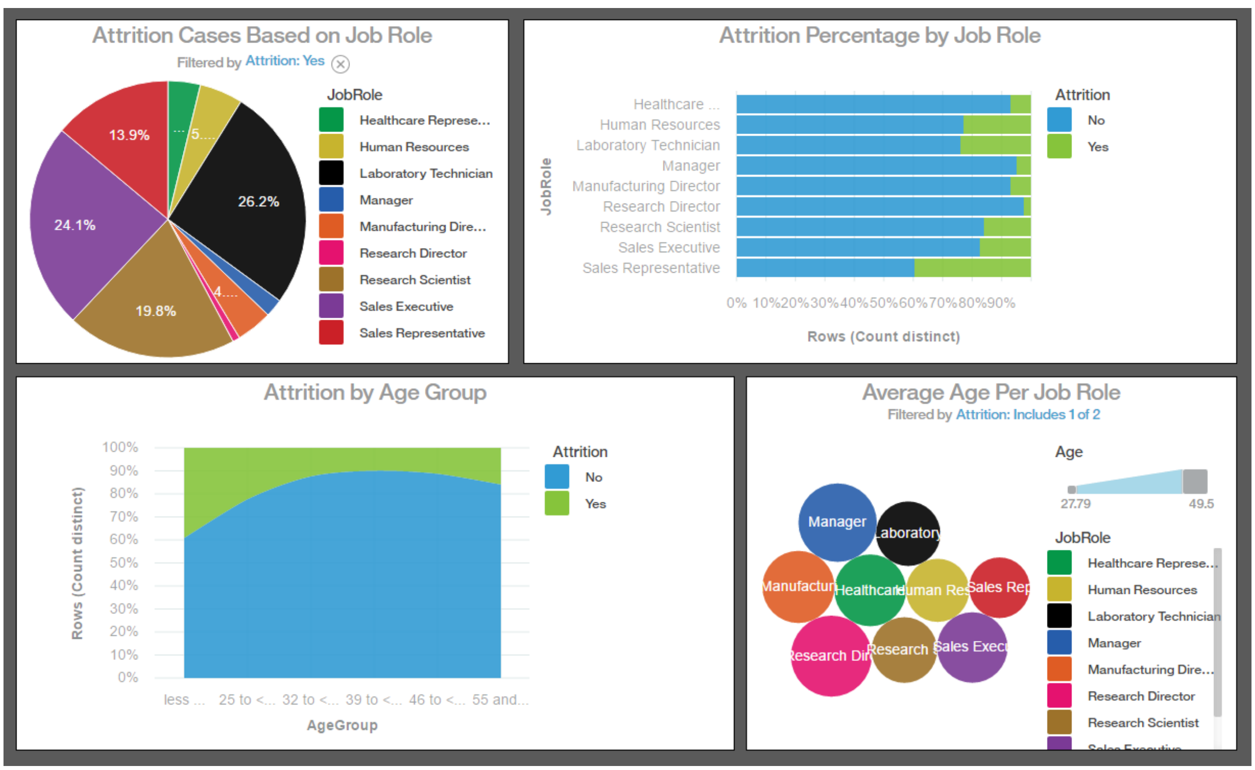

Figure 1. Dashboard of Employee Attrition Based on Job Role and Age.

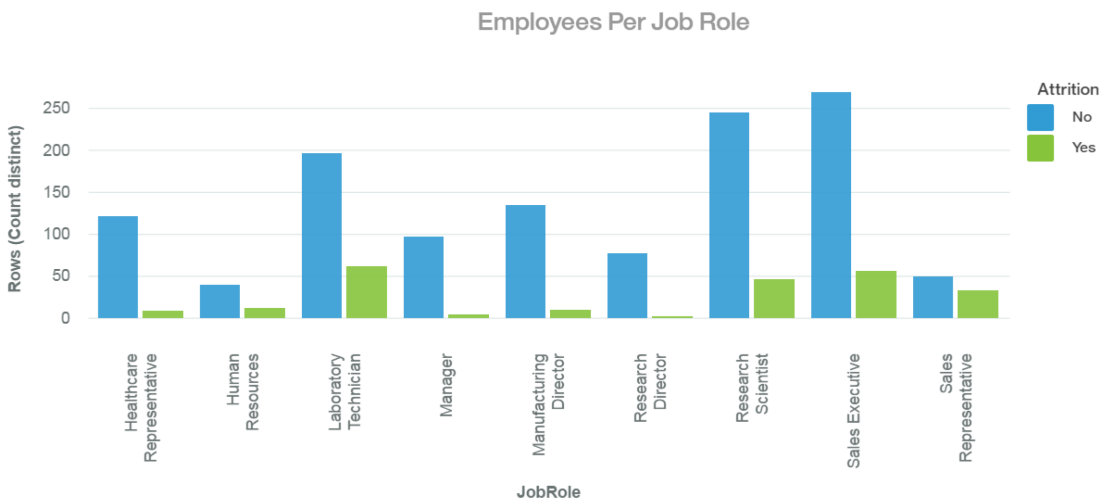

Figure 2. Number of Employees Per Job Role Based on Attrition.

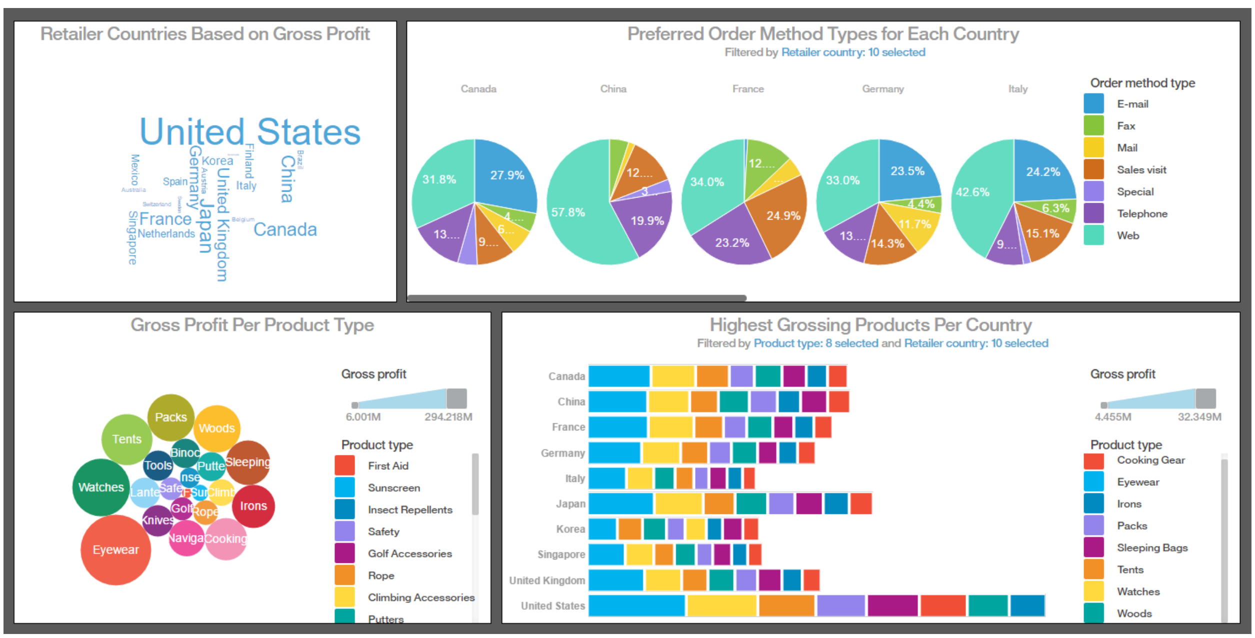

Figure 3. Dashboard of Gross Profit Based on Country, Order Method, and Product Type.

Figure 4. Dashboard of Twitter Tweets Based on Product Hashtag and Author Country.

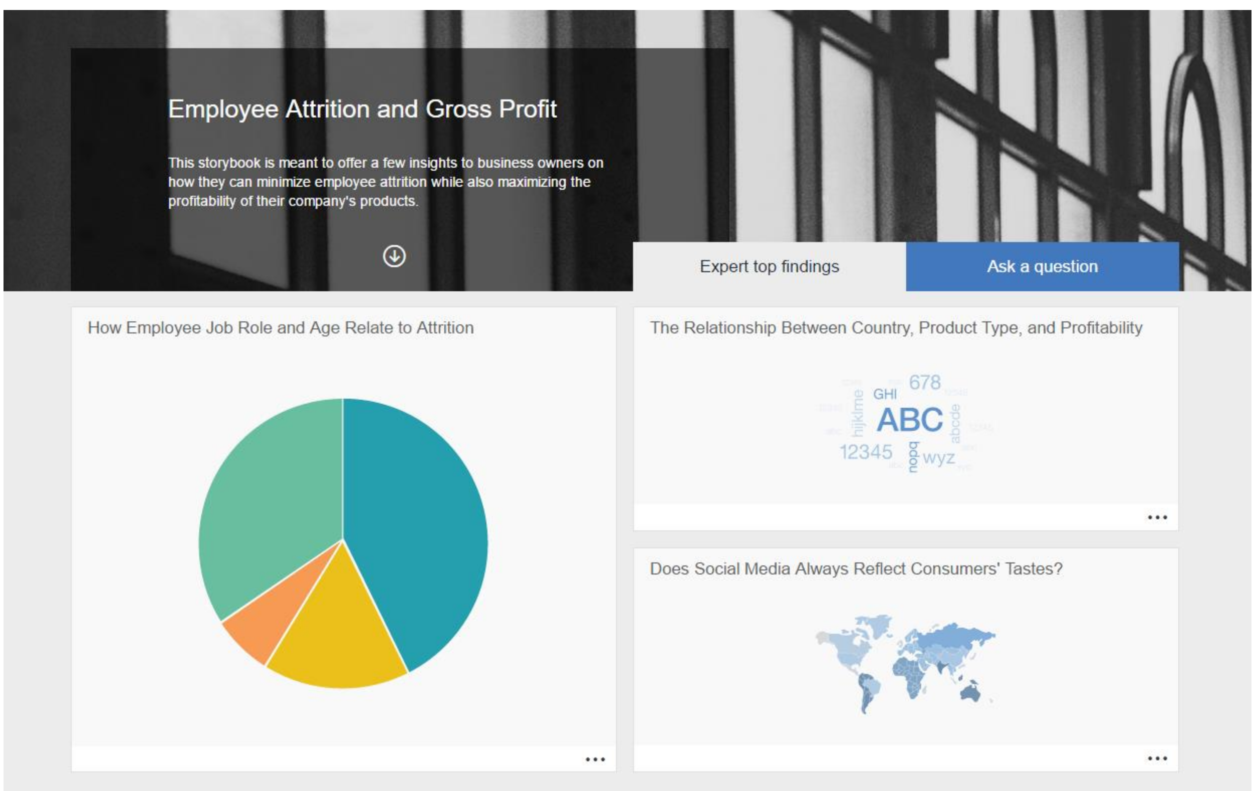

Figure 5. Employee Attrition and Gross Profit Storybook.

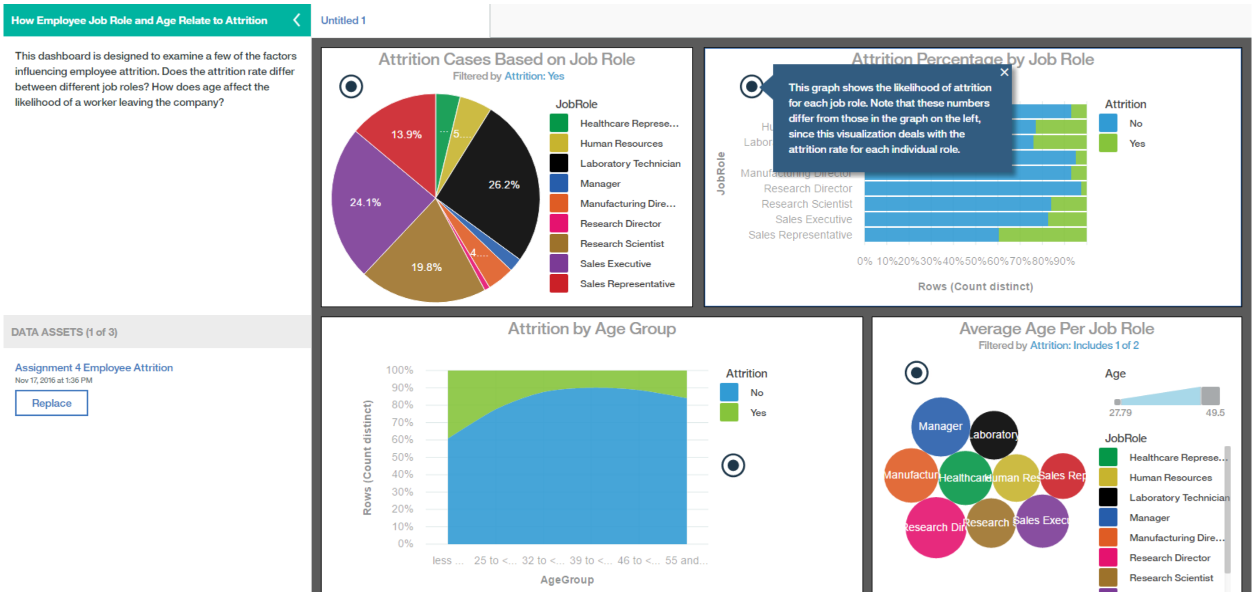

Figure 6. Employee Attrition Storybook Page.

Figure 7. Gross Profit Storybook Page.

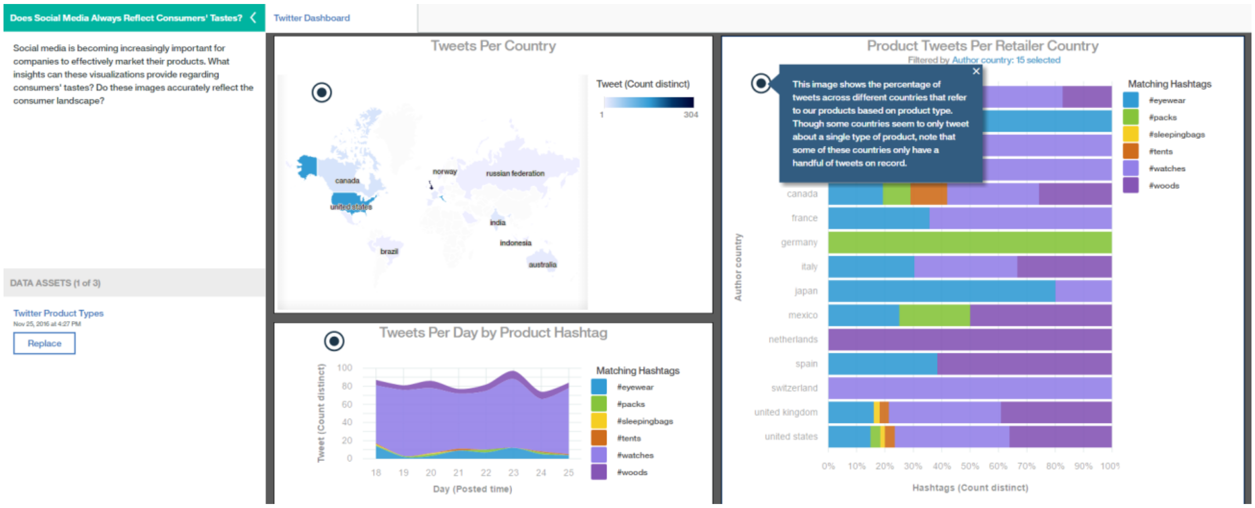

Figure 8. Social Media Storybook Page.

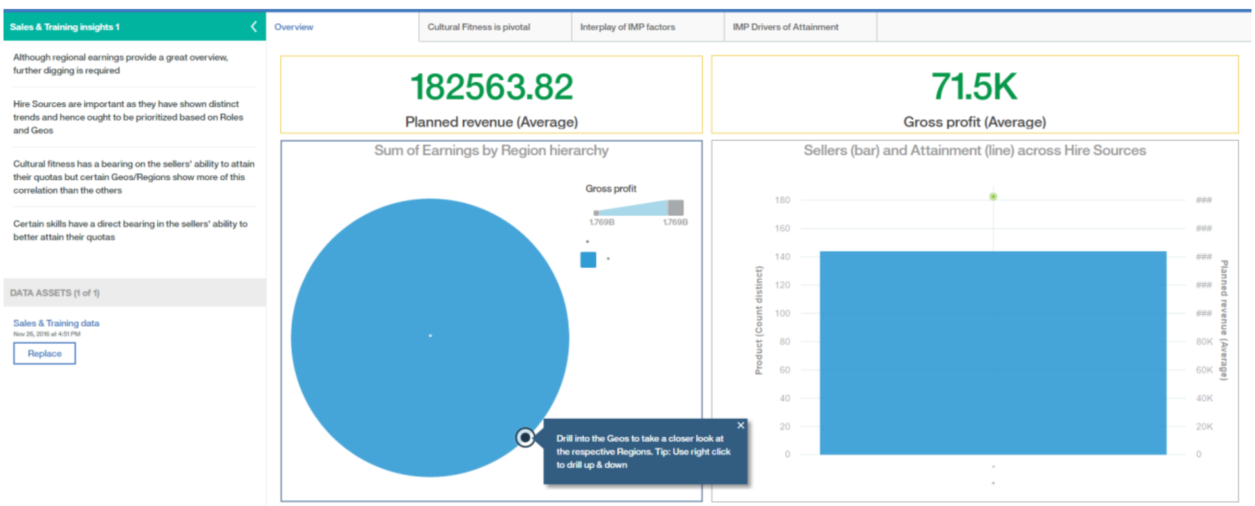

Figure 9. Imported Sales and Training Storybook Page with User Dataset.