Yasin Hassan

Power BI Report

A company has contacted me recently for advertising. They want me to research which product will best fit the specific customer. The products being advertised were “shirts, sweaters and leather bags” (Overview, n.d.). In order to do this, I would need to find where “greatest income exists around the country and whether there is a correlation between sales and income” (Instructions, n.d.). I will need to answer five questions in order to find this information:

* What is the correlation (R2 value) between sales and income?
* What is the linear regression formula to predict customer sales and customer incomes?
* What is the correlation (R2 value) between customer ratings and product return rate?
* Which customer do you predict has the highest income?
* Which product will be advertised the most?

I needed to find the correlation and linear regression formula between sales and income. I found the average sales for 6 months for each state as well as the average income for each state. I created columns for these measurements in my regression table. I created a scatterplot with the average sales as the independent variable(x) and the average income as the dependent variable(y). As the average sales went up, the income went up as well. The correlation for the average sales and average income is 0.78.

In order to find the linear regression formula, I needed to find m and b, the y-intercept. First I had to find n, the number of columns in the regression table. Then, I created columns for x^2, y^2, and xy. Afterwards, I had to find the measurements sum\_x, sum\_y, sum\_x^2, sum\_y^2, and sum\_y. To find b, I used the formula “b = DIVIDE([Sum\_Y]\*[Sum\_X^2]-[Sum\_X]\*[Sum\_XY],[n]\*[Sum\_X^2]-[Sum\_X]^2)” (Linear Regression Part 2, n.d.) . To find m, I used the formula “m = DIVIDE([n]\*[Sum\_XY]-[Sum\_X]\*[Sum\_Y],[n]\*[Sum\_X^2]-[Sum\_X]^2)” (Linear Regression Part 2, n.d). The final formula I got was y = 72.43x + 72638.21

To find the correlation value between customer ratings and product return rating, I took similar steps in finding the previous correlation value. However, this time, I didn’t make any new columns. I did have to split the Purchase List table with one column into six different columns. There was also inconsistent spacing in the columns, so I had to fix that issue as well. One I did, I created another scatterplot, using customer ratings as the independent value and the return rate as the dependent value. As the customer rating increased, the return rate decreased. This makes sense because if customers like a product more, then there will be less customers returning it. The correlation value for the customer ratings and return rate is 0.69.

I needed to find the customer with the predicted highest income. I decided to use my linear regression formula earlier in the Customer List Table. I plugged in the [Last 6 Months Purchases] as the x value, and used the same m and b. After I sorted the column using the descending option, I found John Little from Illinois (JLit30836) has the highest income.

Now I need to figure out which product will be advertised the most. I had to advertise shirts, sweaters, and leather bags to the correct customers. In order to figure out which products should be advertised to who, I created bins for the income ranges of each customer. I created a histogram using these bars, where the height of each bar is the count of customers in a certain income range. I also made some rules regarding advertising the products.

* Since shirts are $25, I will advertise it to customers whose incomes are greater than or equal to 0 and less than $60,000.
* Since sweaters are $100, I will advertise it to customers whose incomes are greater than or equal to $60,000 and less than $120,000.
* Since leather bags are $1000, I will advertise it to customers whose incomes are greater than or equal to $120,000.

I created a bar graph comparing which item is most popular out of the three products and shirts seem to be the most popular product. This makes sense because most customers have incomes in between $0-$60,000. This means that we should advertise shirts the most.

One topic I wanted to figure out for additional research is the population for each state. This is important because the higher the population for a state is, the more customers you have. Also, companies can decide if the population for a state is high enough to set up business there. I made a bar graph that counts the population for each state. The bar graph is also cross-filtered with a bubble map, where if you click on a bar on the bar graph, which represents a state, the map will zoom in to the specific state, where you can hover over the bubble and view the average income for the state. If we focus on the rules mentioned earlier with the shirts, sweaters and leather bags, we can use the average income for each state to figure out which product to advertise the most in each state. For example, California has the highest population and the average income is 106269.18. This means that we want to advertise sweaters the most in California.

References

*Calculated Columns.* (2020, October 10). Sqlbi. Retrieved November 4, 2021, from

https://www.sqlbi.com/articles/calculated-columns-and-measures-in-dax/

*Correct Citation.* (n.d.). Purdue Owl. Retrieved November 4, 2021, from

<https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_g>

uide/reference\_list\_electronic\_sources.html

*Count Rows*. (2021, November 3). Udacity. Retrieved November 3, 2021, from

<https://knowledge.udacity.com/questions/733337>

*Cross Filter.* (2021, November

*Duplicate Error*. (2021, November 2). Udacity. Retrieved November 2, 2021, from

<https://knowledge.udacity.com/questions/733123>

*Histogram*. (n.d.).Udacity. Retrieved November 2, 2021 from

[https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f6](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/94de46d4-8137-47e7-8e7c-3653de6dc767)

[5ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/94de46d4-8137-47e7-8e7c-3653de6dc767)

[320-8637-e2e710b3edb0/concepts/94de46d4-8137-47e7-8e7c-3653de6dc767](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/94de46d4-8137-47e7-8e7c-3653de6dc767)

*Images*. (n.d.). Udacity. Retrieved November 4, 2021, from

<https://knowledge.udacity.com/questions/692459>

*In-Text Citations*. (n.d.). Purdue Owl. Retrieved November 4, 2021, from

<https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_g>

uide/in\_text\_citations\_author\_authors.html

*Linear Regression Part 1*. (n.d.). Udacity. Retrieved November 1, 2021, from

[https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f6](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/ebcdc02b-c863-468f-a69f-6418c67f9e99)

[5ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/ebcdc02b-c863-468f-a69f-6418c67f9e99)

[320-8637-e2e710b3edb0/concepts/ebcdc02b-c863-468f-a69f-6418c67f9e99](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/ebcdc02b-c863-468f-a69f-6418c67f9e99)

*Linear Regression Part 2*. (n.d.). Udacity. Retrieved November 1, 2021, from

[https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f6](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/a02a4ad1-f4b2-41bd-879e-35511d5bade1)

[5ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/a02a4ad1-f4b2-41bd-879e-35511d5bade1)

[320-8637-e2e710b3edb0/concepts/a02a4ad1-f4b2-41bd-879e-35511d5bade1](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/a02a4ad1-f4b2-41bd-879e-35511d5bade1)

*Predicted Income*. (2021, November 2). Udacity. Retrieved November 2, 2021, from

<https://knowledge.udacity.com/questions/732449>

*Recommended Product*. (2021, November 4). Udacity. Retrieved November 4, 2021, from

https://knowledge.udacity.com/questions/734604#734665

*Regression Formula*. (2021, November 2). Udacity. Retrieved November 2, 2021, from

<https://knowledge.udacity.com/questions/732449>

*Salary Bin*. (n.d.). Udacity. Retrieved November 3, 2021, from

[https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f6](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/d9e7d43a-04c7-491f-bf1d-92d3a62dab2e)

[5ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/d9e7d43a-04c7-491f-bf1d-92d3a62dab2e)

[320-8637-e2e710b3edb0/concepts/d9e7d43a-04c7-491f-bf1d-92d3a62dab2e](https://classroom.udacity.com/nanodegrees/nd331/parts/de126807-159c-47df-885d-f9f65ee9fedd/modules/f6679eaa-38cb-4832-a6eb-cb5fe21b6942/lessons/ca61d10d-e537-4320-8637-e2e710b3edb0/concepts/d9e7d43a-04c7-491f-bf1d-92d3a62dab2e)

*Specific Locations*. (2021, November 4, 2021). Udacity. Retrieved November 4, 2021, from

https://knowledge.udacity.com/questions/734908

*Summarize Columns*. (2021, October 25). Udacity. Retrieved November 1, 2021, from

<https://knowledge.udacity.com/questions/720978>