From Udacity

Program: Data analysis and visualization with Power BI

**Project description**

A national online clothing chain needs your help to create a targeted marketing campaign. Sales have been poor and they want to attract lost customers. They want to advertise specific products to specific customers in specific locations, but don't know who to target. They have three products in mind:

* Shirt: $25
* Jumper: $100
* Leather bag: $1,000

They need you to do an analysis to determine the best product to advertise to each customer.

**Resources**

All project work will be done with the Microsoft Power BI Desktop application.

To view the raw spreadsheet data outside of that platform, students will need to use Google Sheets or Microsoft Excel.

The project will use several data sources, including:

**US Census Bureau**

* Median income
* Location
* Population
* Sector

**Company data**

* Inventory of products
* Product prices
* Customer ratings
* Product return rate

**Customer data**

* Customer identification
* Names
* Location
* Date of birth
* Purchase history

**Additional data**

* Meteorological characteristics
* Economic data
* Demographic data
* Competition

**Project Instructions**

In this project, you will use population statistics from the US Census Bureau to determine where the highest incomes are located across the country and whether there is a correlation between sales and income. We don't know our customers' incomes, but we should be able to predict them by looking at their purchase history and location, and comparing it to census data. In addition, we want to analyse our inventory, specifically customer ratings and return rates, and see if there is a correlation between the two.

**Draw conclusions**

Draw conclusions from your analysis and use images to answer the following questions:

**Analysis questions:**

* What is the correlation (R2 value) between sales and returns?
* What is the correlation (R2 value) between customer ratings and product return rate?
* What are the linear regression formulas for predicting customers' revenues from their sales?
* Which customer do you think has the most revenue?
* Which product will be advertised the most?

**Present your analysis**

You will need to present your analysis as a 1-page written summary and a visual report in Power BI. Use the following images to present your data:

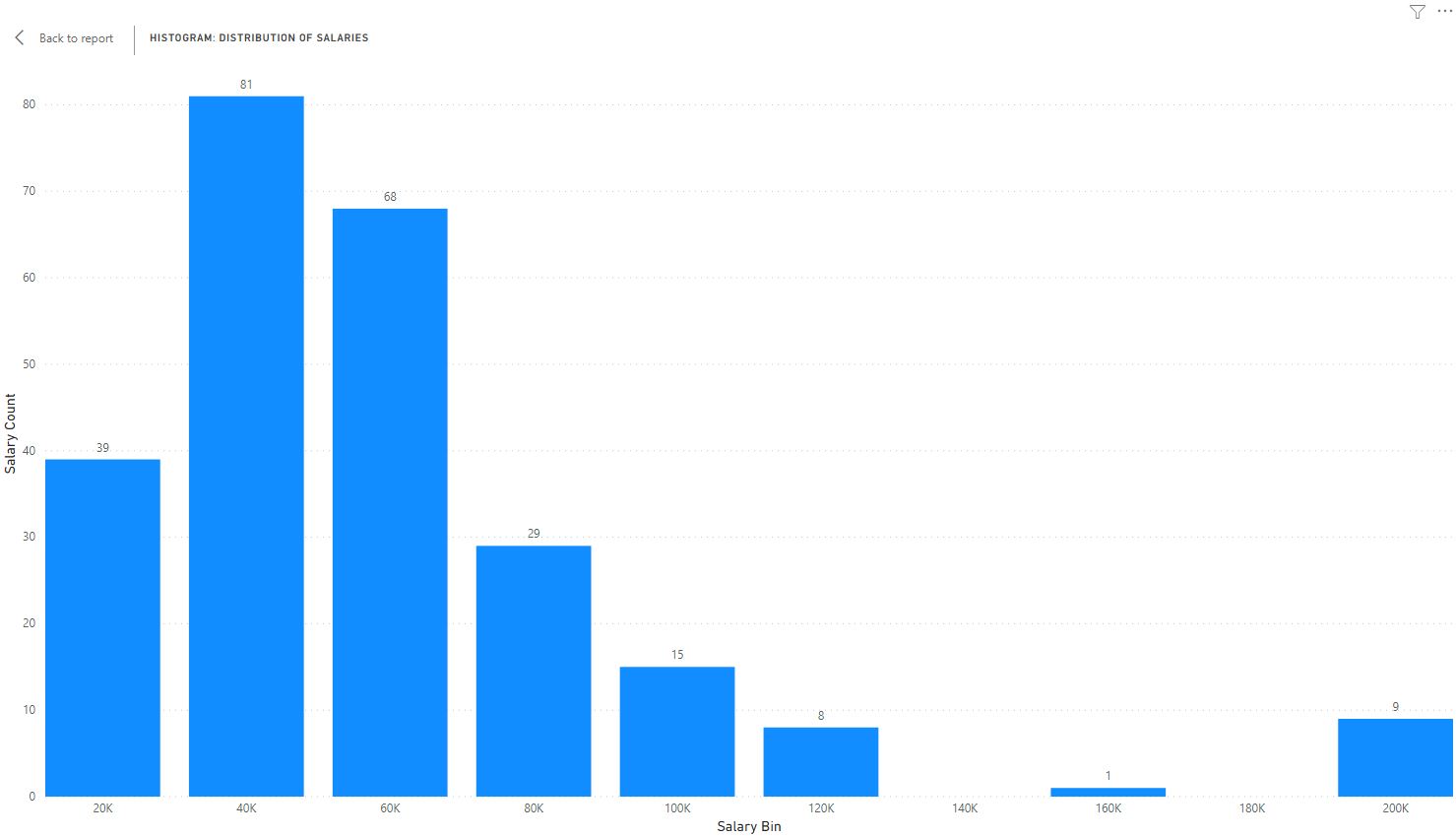
* Income distribution Histogram
* Household income by location: Map
* Correlations: Scatter plot with trend line and R^2 value card
* Use other visual objects as needed to better present the results of your analysis.

**Histogram**

Although you will need to determine the parameters of the histogram(s) for your project, think about the purpose of a histogram. In the example below, you can see that the histogram allows you to clearly understand the distribution of values and helps you visualise the shape of the data. We can see here that this is a distribution skewed to the right, with a peak of around $40,000 and a mean that is probably a little to the right of the peak.

Your histogram(s) should also convey important statistics about the data you are analysing, so you may want to have at least 4 columns, although it is preferable to have a few more.

Also, although there is a visual histogram visual object that you can download from the Power BI Marketplace (if you have access to that resource), it is recommended that you use a column chart instead so that you can use a DAX formula and specify the ranges and number of columns (you will find more information on this in the detailed instructions).



**Additional analysis**

Finally, choose one or two more variables to analyse further and include in your report. The variables can come from existing data sets or you can choose them from another source. Think about how the data relates to your revenue forecasts and sales recommendations while helping you better understand customer demographics and market conditions, allowing you to create a more rigorous analysis.