 [My Programs](https://learn.udacity.com/)

 [Project: Market Analysis Report for National Clothing Chain](https://learn.udacity.com/nanodegrees/nd331-sc/parts/7b38acf7-fcf8-4d73-bb3e-d6e56d67f9e2/lessons/1107eff7-a331-4b8e-817b-3885af13890e)

1. Project Overview

**Project Overview**

**Market Analysis Report for National Clothing Chain**

|  |
| --- |
| Shopping mall imageinside clothing store image |

**Project Description**

**An online national clothing chain needs your help creating a targeted marketing campaign. Sales have been flat and they want to lure lost customers back. They want to advertise specific products to specific customers in specific locations, but they don’t know who to target. They have three products in mind:**

* Shirt: $25
* Sweater: $100
* Leather Bag: $1,000

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

Image Information

* "fashionistas" by [ embr ] is licensed under CC BY-NC-SA 2.0
* "French Connection @ Shopper's Stop, Garuda Mall" by teemus is licensed under CC BY-NC-SA 2.0

**Environments and Data**

**Project Environment and Data Sources**

All work for the project will be completed using the Microsoft Power BI desktop application.

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

The project will use a variety of data sources, including

**US Census Bureau**

* Average income
* location
* population
* industry

**Business Data**

* Product inventory
* Product prices
* Customer rating
* Product return rate

**Customer Data**

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

**Additional Data**

* Weather
* Economics
* Demographics
* Competition

**Instructions Overview**

**Project Instructions**

**Project Instructions**

In this project, you will use population statistics from the US Census Bureau to determine where the greatest income exists around the country and whether there is a correlation between sales and income. We don’t know the incomes of our customers, but we should be able to predict it by looking at their purchase history and locations and comparing that against the census data. Additionally, we want to analyze our inventory, specifically customer ratings and return rate and see if there’s a correlation between the two.

**Draw conclusions**

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

Analysis Questions:

1. What is the correlation (R2 value) between sales and income?
2. What is the correlation (R2 value) between customer ratings and product return rate?
3. What are the linear regression formulas to predict customer sales and customer incomes?
4. Which customer do you predict has the highest income?
5. Which product will be advertised the most?

**Present your analysis**

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

* Income Distribution: Histogram
* Household Income by Location: Map
* Correlations: Scatter Plot with Trendline and Card with R^2 value

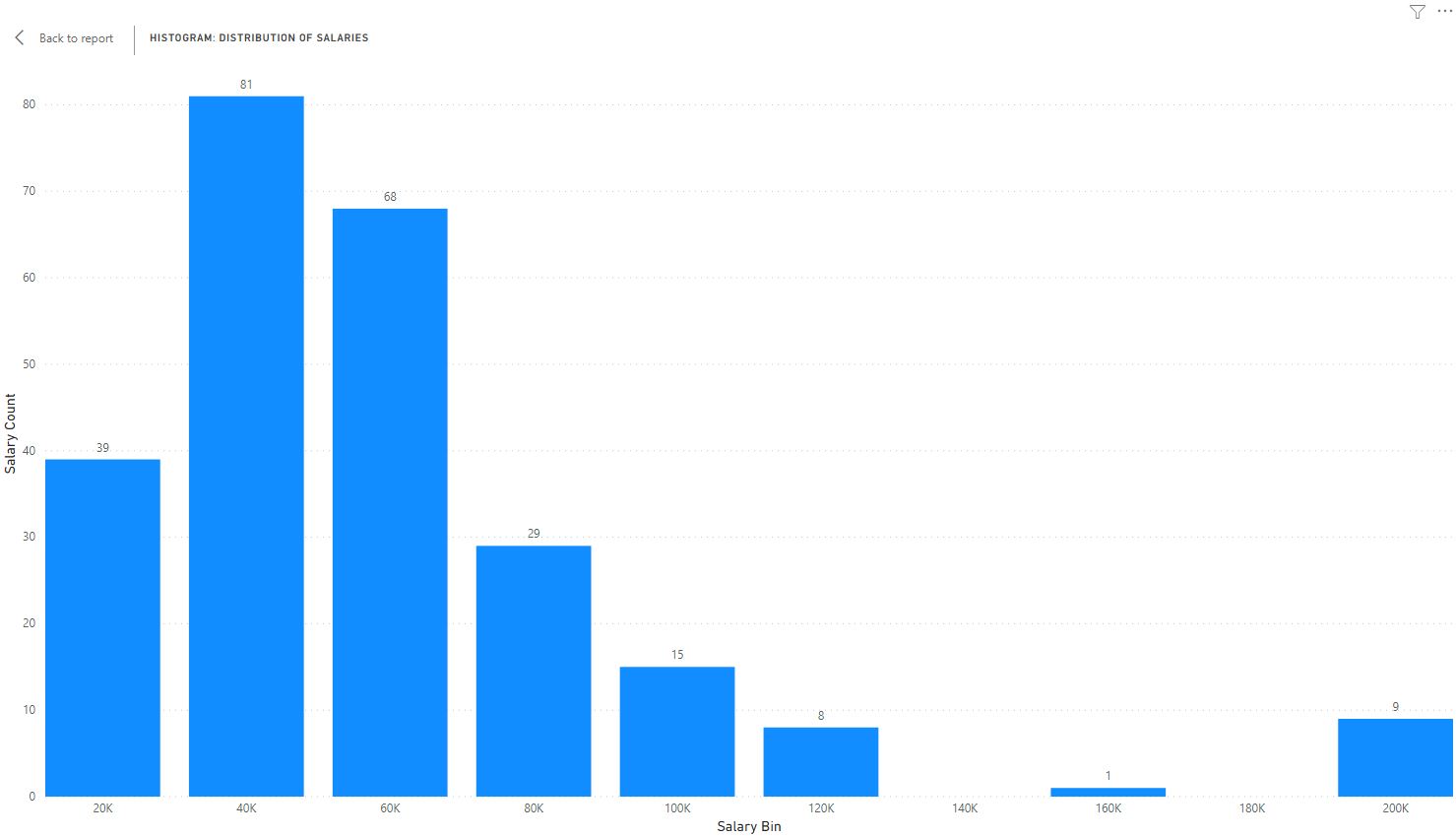
Use other visuals as-needed to further present the results of your analysis.

**Histogram**

Although you will need to determine the parameters of the histogram(s) for your project, think about what a histogram is intended to do. In the example below, you can see that the histogram gives you a clear understanding of the distribution of values and helps you visualize the shape of the data. We can see here that it's a right-skewed distribution with a peak of around $40k and an average that is probably a bit to the right of the peak.

Your histogram(s) should also convey important statistics about the data you analyze, so you'll probably want to have at minimum 4 columns but preferably more.

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



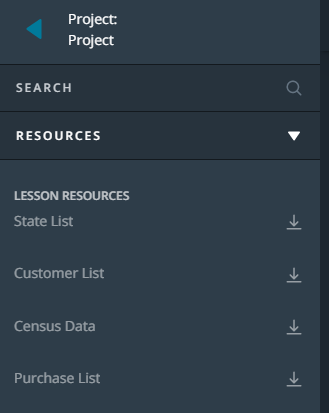
**Further analysis**

Finally, choose 1-2 other variables to analyze further and include in your report. The variables can come from the existing data sets or can be of your choosing from another source. Think about how the data relates to your income predictions and sales recommendations. and may help you better understand the customer demographics and market conditions and that will allow creating a more rigorous analysis.

# Detailed Instructions

## Detailed Instructions

First, you will need to download the 4 spreadsheets needed to complete the project. These are located under the Resources section of the page in the upper left corner of this page (see screenshot below)



Files needed for the project located top left of page

Detailed Instructions

1. Import the census data, customer list, purchase list, and state list into Power BI. You’ll use the “Get Data” button to start the import process. Make sure you select each of the available tables within each Excel file. There should be 7 total tables:
   * Avg Income by State
   * Customer List
   * Incomes by State
   * Industries
   * Product Inventory
   * Purchase List
   * State List
2. Set up the data in Power Query so that all the columns are correctly formatted and structured. Some of the tables will require steps in Power Query to be correctly set up.
3. Set up your table relationships. All tables you import should be tied into the table relationships. Avoid many-to-many relationships as they’re not needed and will cause issues with cross-filtering.
4. Review the table relationships and confirm if they are properly set up. Most of the tables should follow a star schema and the overall set up should look similar to the sample image:

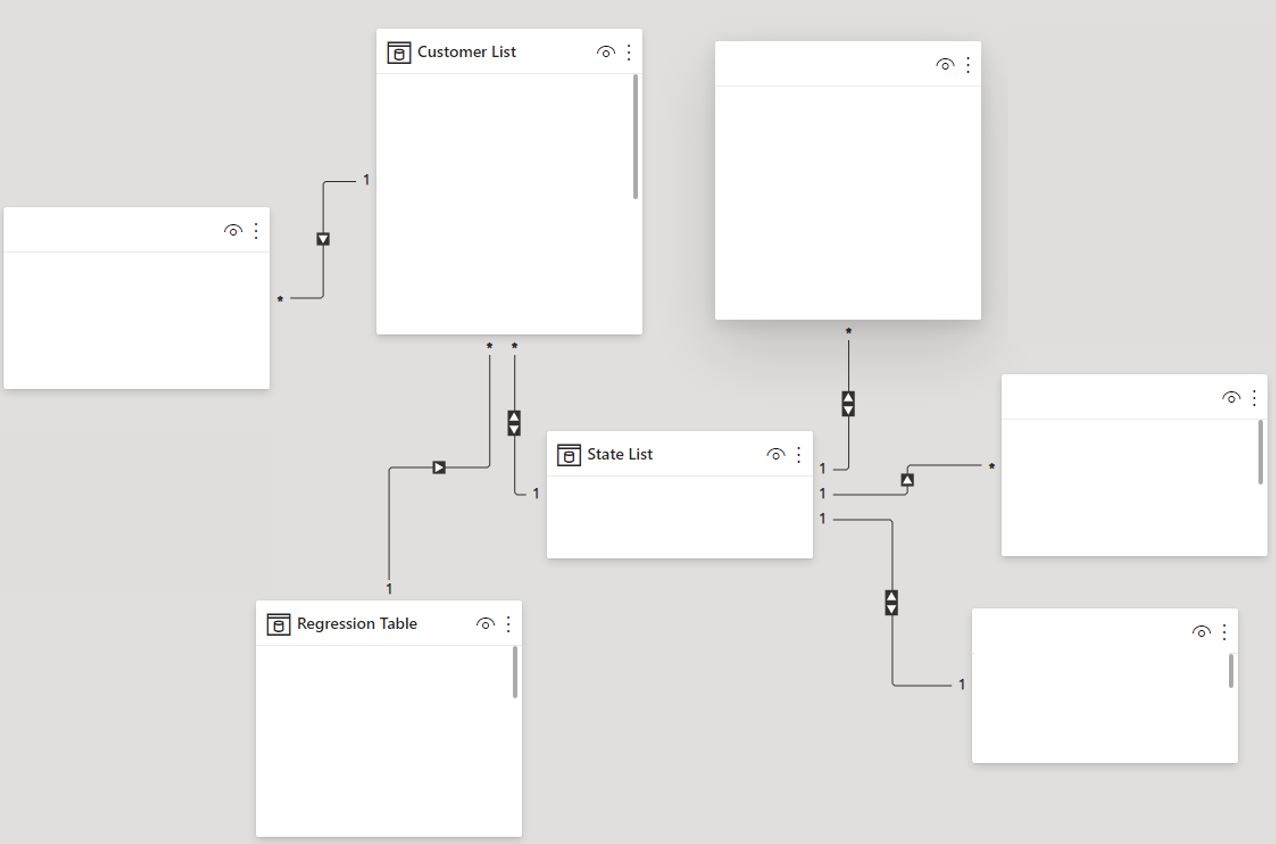


Table Relationships diagram sample

5.Create a map to visualize the distribution of average household income across the country.

6.Create a new ‘Regression Table’ from the ‘Customer List’ to summarize the average customer sales by state and set up the following calculated columns for your regression formula:

* y = average sales by state (read the note below)
* y2 = Y^2
* x = average income by state (read the note below)
* x2 = X^2
* xy = X\*Y

NOTE: You'll need to use this data to calculate predicted customer incomes, and this can be set up in 1 of 2 ways. Either (1) set up income as x and use x = b-y/-m to predict customer income, or (2) set up sales as x and use y = mx +b to predict customer income. The results will be very similar.

7.Create a scatterplot with a trendline to analyze the relationship between average incomes and average sales.

8.Create the remainder of the regression formula variables as calculated measures: b (y-intercept), m (slope), n, Sum of X, Sum of Y, Sum of XY, Sum of X2, Sum of Y2

9.Create a calculated column in your customer table for your predicted customer incomes.

10.Use DAX formulas to categorize the predicted incomes into ranges and to determine the best product fit for each customer. Present this data with a histogram and other visuals.

11.Perform additional analysis using 1-2 other variables of your choosing. These variables can be part of the provided data set or data that you find from additional sources. Think about variables that may help you better understand the customer demographics and market conditions and that will allow creating a more rigorous analysis.

12.Write a brief 1 page summary of your findings, citing the visuals and tables you created as evidence. The summary should address the 5 analysis questions described on the **Instructions Overview** page and provide an overview of your resulting marketing strategy.

Please use the [rubric (opens in a new tab)](https://learn.udacity.com/rubric/3078)to verify your work

# Project Submission Details

## Final Submission

### Submission

As previously mentioned, you'll have at least two files to submit, but maybe a third if you incorporate additional data sources into your analysis.

* 1 - 2 page summary report in a word processor document
* Your completed Power BI file
* (if you choose to use additional data from other sources), submit those files as well

After you submit your project, it'll be reviewed by an actual human and you'll be given real feedback on your work along with any questions or further instructions if needed.

# Project: Market Analysis Report for National Clothing Chain

## Power Query

| **Success Criteria** | **Specifications** |
| --- | --- |
| SWBAT organize, clean and format data using the pre-set Power Query tools. | The joined data on the ‘Product Inventory’ table of the Customer List is split into 6 columns, each labeled with correct formatting and no resulting Power Query errors.  Submission: Screenshot of final table in Power Query (include applied steps)  Submission: Also submit a final copy of your Power BI file along with all screenshots. |
| SWBAT organize clean and format data by creating custom M formulas. | The ‘Purchase List’ table is un-pivoted, organized, and has a date column that is correctly formatted as a date. There should be no resulting Power Query errors for any of the columns or rows.  Submission: Screenshot of the final table in Power Query (include applied steps)  Submission: Also submit a final copy of your Power BI file along with all screenshots. |

## DAX

| **Success Criteria** | **Specifications** |
| --- | --- |
| SWBAT create a calculated column with DAX logic to generate a histogram | The income categories should be defined using a DAX formula. The DAX formula should aggregate the different predicted customer incomes into buckets which can be used to create a histogram. The appropriate bin size for the histogram can be determined by the student but should still be a good reflection of the range, distribution, and shape of the data. It is recommended that the histogram contain at least 4 columns. Refer to the histogram example in the instructions section.  Submission: Screenshot of DAX formula Submission: Also submit a final copy of your Power BI file along with the screenshots. |
| SWBAT create a calculated column or measure with DAX logic to generate a column chart | The product recommendations should be defined using a DAX formula. The DAX formula should use logic to determine which products are recommended to different income categories. The recommended product for each income category can be determined by the student.  Submission: Screenshot of DAX formula  Submission: Also submit a final copy of your Power BI file along with all screenshots. |

## Visualization & Analysis

| **Success Criteria** | **Specifications** |
| --- | --- |
| SWBAT use linear regression to predict future outcomes | A formula is created that can be used to predict customer incomes based linear regression of sales and income. Using y = mx + b, the m and b variables are replaced with the actual values and presented in the written summary. • Submission: The formula is included in the written summary and Power BI file. • Submission: Screenshot of the formula found in the Power BI file. • Submission: Also submit a final copy of your Power BI file along with all screenshots. |
| SWBAT analyze data with histogram visualization | The histogram shows the distribution and shape of predicted income by category. The histogram is created using a column chart and DAX formula (the calculated column created earlier) to define the ranges/bins of the columns.  Submission: Screenshot of histogram Submission: Also submit a final copy of your Power BI file along with the screenshots. |
| SWBAT analyze data with scatter plot and card visualization | The scatter plot with trendline and correlation coefficient quick measure (on a card) is used to perform a regression analysis of the relationship between average household income by state and average 6 months sales by state.  Submission: Screenshot of scatterplot and card  Submission: Also submit a final copy of your Power BI file along with all screenshots. |
| SWBAT analyze data with heatmap visualization | The heatmap is used to visualize income household income distribution across the US.  Submission: Screenshot of the heatmap  Submission: Also submit a final copy of your Power BI file along with all screenshots. |
| SWBAT set up table relationships so that visualizations correctly cross filter | Cross-filter: The histogram columns can be used to update the heatmap. The scatterplot can be used to update the histogram. The scatterplot can be used to update the map.  Submission: Screenshot of cross-filtered histogram and heatmap Screenshot of cross-filtered scatterplot and histogram Screenshot of cross-filtered scatterplot and heatmap  Submission: Also submit a final copy of your Power BI file along with the screenshots. |
| SWBAT present the findings of their analysis and as conclusive, evidence-based recommendations. | The 1-2 page written report provides a detailed summary of the results, conclusions and recommendations of the analysis. The document reads like a well-written executive summary and includes the following: • All 5 of the analysis questions are addressed with 1-2 sentences • The formula for predicting customer incomes • The scatterplot relationships with R-Squared values • Findings from the research of 1-2 additional variables (as noted in the instructions) are included • Any findings that are used to inform the marketing strategy • Specific visuals that speak to the narrative of summary are included in the summary. • Final recommendations are presented and are based on the results of the statistical analysis.  Submission: A 1-2 page document (a little longer is ok) |

### Suggestions to Make Your Project Stand Out

1. Further analyze the data sets. Not all relationships, correlations, or distributions have been explored. Think about these questions:

* Is there a way you could recommend specific clothing styles by region?
* Is there a way to explain the sales trends over time? What is the variance or standard deviation?
* Are there any states that have a disproportionately large customer base when compared to the population? How can that information be leveraged?

1. Bring in outside data to add further context to the analysis. Think about how these variables could be brought in and used to more precisely drive a marketing strategy:

* Weather (precipitation, temperature, etc):
  + National Oceanic and Atmospheric Administration NOAA: noaa.gov
* Economic Data (unemployment rate, industries, credit ratings, etc.):
  + Bureau of Labor Statistics: bls.gov
  + Federal Reserve: fred.stlouisfed.org
* Population Data (age, ethnicity, education, etc.)
* Census Bureau: data.census.gov/cedsci
* Competition Data
  + What national or regional retailers are considered competitors and why?
  + How do competitors' product offerings and prices compare?
  + Are there locations with minimal or no competition that could present as good opportunities to expand?

1. Consider using other advanced visuals covered in this course to present your data:

* KPI
* Decomposition Tree
* Funnel
* Radial Gauge
* Waterfall

**Project: Project: Market Analysis Report for National Clothing Chain**

Past Due on March 29, 2024

The final project you submit should be a complete and thorough market analysis in Power BI and 1-page summary write-up using a word processor program. The relationships between the different tables, such as age, income, location, etc should be presented in visuals that are clear and engaging. The findings of your statistical analysis should be explained in the 1-page document, but also evident in the Power BI file with proper labels and color coordination. There should be no formula errors either in DAX or Power Query. Most importantly explain your marketing strategy using data to back up that strategy.

The final zip file should include the following

1. PowerBI file downloaded from the website
2. Analysis summary document

The zip file should be called, "Final Project"

**How would you like to submit your project?**

Select Submission Type