CAPSTONE  
  
  
ALY6980, SPRING 2021  
MODULE 8 PROJECT ASSIGNMENT

WEEK 8: THE INDIVIDUAL CAPSTONE PROJECT

SUBMITTED BY: SHIVANI ADSAR

NUID: 001399374

SUBMITTED TO: DR. ERIC (YANG) LIU

DATE: 06/10/2021

**Introduction**

The assignment aims at performing an analytical project proposal that will provide in-depth insights about the work done for data analysis on At Home Group’s data. Moreover, this assignment will focus more on the purpose of the project, background research, data governance, implementation methodologies, and data collection methods. Moreover, the assignment will highlight the importance of data governance while implementing data analysis for the company. This analysis will provide insights about the data governance techniques and methods to solve the business problem and ways to provide optimal solutions to the client through improved business recommendations and strategies.

**Analysis**

Statement of Purpose

The project contains data about the At Home Group, which is a furnishing home décor industry. We have performed data analysis on the At Home Group data to provide business recommendations to the company for improving their revenue. Since the Capstone course involves working with a company on their data for data analysis, we have collaborated with At Home Group, through the university. The data has been gathered and collected through initial market research from various online websites. The data consisted of ethnic, demographic, regional, and tax level zip code data. This data was collected for 227 store locations present in the US in various states. The business question is about improvising the budget for the At Home Group stores present in the United States, and ways to improve sales through improvised marketing techniques, by performing data analysis on the company’s data. Moreover, the business questions would involve providing business recommendations based on the data analysis. The problem to be solved for the business question is the identification of higher customer sales at a given location, and then, tracking the growth of a store at a particular location based on zip code level data. The research question for the data is about using the researched demographical, tax, and regional data of the 227 At Home Group stores for performing store-level business analysis. The goals for the project proposal are:

* Improve the sales for the stores
* Improvise the store’s e-commerce websites
* Proposing business recommendations based on data analysis
* Propose marketing techniques by studying customer buying patterns

These goals and business recommendations will help the organization in analyzing ways to implement various business strategies to improve overall business outcomes. Moreover, the data visualizations will provide great data insights about the dependency of a parameter on the other and understand the directly proportional parameters.

Scope of Project

The scope of work is given by the milestones to be achieved within a given period. Some of the areas of work include:

* Data Research: This involves researching the data before performing data gathering
* Data Collection: This involves collecting the data based on the researched data
* Data Analysis: This involves performing exploratory data analysis and then using the data insights to provide a comparative analysis of sales. Moreover, this would also involve the implementation of further data analysis algorithms like K-means, KNN, Linear Regression, etc. for understanding the dependency amongst variables.
* Business Recommendations: These will be provided based on the data analysis done earlier, and will help improve the company’s revenue.

The project involves some key deliverables as follows:

* Document for summarizing the research performed on the home décor industry
* Excel document, as developed through market research
* Report on the exploratory data analysis, along with Tableau file
* Business Recommendation
* Project Proposal on the project

The timeline for the project is for 12 weeks and will be completed by the 3rd July 2021. Moreover, every deliverable has a timeline of 14 days.

Background Research and Literature

As we know, it is important to perform background research before performing data analysis. It is important to maintain data governance to ensure data integrity and security, especially while working with technical data. This helps in maintaining discipline within the company data. It is a good point to implement data governance within the project as the data will remain consistent and ensures critical data is conveyed to the correct individual within the organization. Therefore, data governance, leadership, and ethics increase efficiency within the organization. The literature that helped in research on data governance is a peer-reviewed article called, Governing Big Data, which tells us about quintillions of data that is being generated every day in the form of media, which needs data governance while working on big data. Moreover, the article talks about managing metadata through data governance as metadata is important for companies to extract analytical insights. In addition, it is important to secure the data to get quality results. Therefore, organizations should focus on implementing centralized data governance measures.

Moreover, there are more literature articles to support the summary of Governing Big Data article. The second article talks about governing big data in financial institutions. The article talks about, strategic approach that should be followed by the financial institutions for managing their data integrity. Moreover, cyber risk is very common in the financial industry, as the consequences of any financial attack or breach can lead to a major loss for the associated people. Therefore, it is important to maintain privacy and ethics for the critical data. Moreover, these companies working on large quantities of data every day should ensure the accuracy, completeness, and timeliness of their data.

Literature:

The first peer-reviewed article is about the IKEA Business Strategy, which talks about how the company uses business strategies for enhancing its sales by focusing on great marketing strategies. Moreover, the company seems to expand by targeting new initiatives. For instance, the company uses Beonic’s Insight Visitor system for tracking if the customer received required staffing which is done through the usage of sensors that track the incoming and outgoing customers within the store. The second peer-reviewed article focuses on ways in which the furniture industry is using new techniques like lean six sigma for optimizing its manufacturing processes. This technique works by reducing the amount of waste and the reduce defects thereby improving the quality of an overall product. This approach has helped small-scale industries working on home décor, to optimize their internal processes and improve sales of products due to their quality. This approach is effective, however, requires skilled workers due to advanced technology. Although the lean six sigma technology has not been implemented widely yet, it is still one of the most effective techniques of saving on wood waste and producing quality products.

The third peer-reviewed article talks more about data governance in financial services and ways to improve the security system to maintain the integrity of data. As we know, financial data is very sensitive, therefore needs constant monitoring. Therefore, financial organizations should focus on keeping an inventory for backing all the data, and then, staying up with the recent advancements in cybersecurity. Moreover, it has been mentioned, that the physical documents should be stored very securely for maintaining their integrity. Also, companies should plan for data breaches in case they encounter them in the future, therefore the companies should be prepared for any such future incidents.

The fourth peer-reviewed article talks about the data governance for healthcare data. As we know (Winter, Jenifer Sunrise, & Davidson, Elizabeth. 2019), healthcare information is very sensitive and should not be disclosed according to some compliance policies. Therefore, healthcare organizations need to abide by HIPPA compliance policies to keep patient data secure.

The fifth peer-reviewed article highlights the importance of data governance while working on huge volumes of data. As we know, huge volumes of data are highly exported and imported into different tools for processing, and it is highly likely for this huge data to get breached by a third-party application. Therefore, organizations should have security trackers within their applications while working with huge data and abide by the data governance rules and ethics.

Design and Data Collection Methods

We have used the data collected from various online resources for working towards the goal of this project. The data consists of regional, demographic, and tax-related details based on the 227 store locations. This data shows a high-level view of zip code level information for every county. The source of this data is government websites that provided accurate and precise zip code level data.

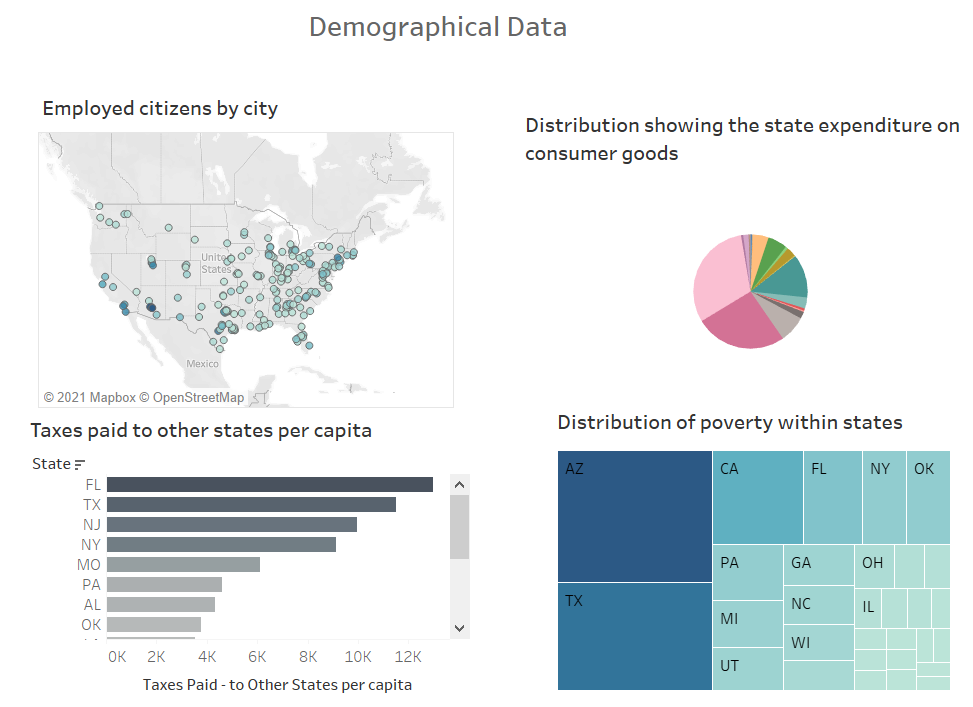
While performing data exploration and analysis, I would propose below tests for maintaining data governance:

* ETL Data Integrity Test: This test will ensure database integrity, thereby maintaining the foreign and primary key integrity.
* BI Verifications: We know that BI tools interact with the database for reporting and dashboarding requirements. Therefore it is important to maintain and plan data integrity tests for these reports.

Implementation Methodology and Strategies

To implement the proposal, I would utilize the implemented exploratory data analysis outcomes to implement the K-Means clustering algorithm. This algorithm will work by performing cluster analysis on the population data which will help in determining the customer buying patterns based on the clusters formed. Moreover, I intend on performing regression analysis which will help in determining the independent and dependent variables and their relationship. This analysis can be used for understanding the parameters that are dependent on sales, which will help in improving the overall business. The data governance will influence the data implementation in a way that, we will have to maintain the data governance and security while working on big data. We will have to gather data from verified sources to avoid any inconsistency in the data. Also, it is important to import the data in tools that have a secure base and are not public. For example, One should avoid the usage of Tableau Public while working on sensitive data as this data gets published, whereas, Tableau Desktop will ensure the safety and security of data. The proposed dashboard is very interactive. The method and strategies that are required to be used while implanting the dashboard are :

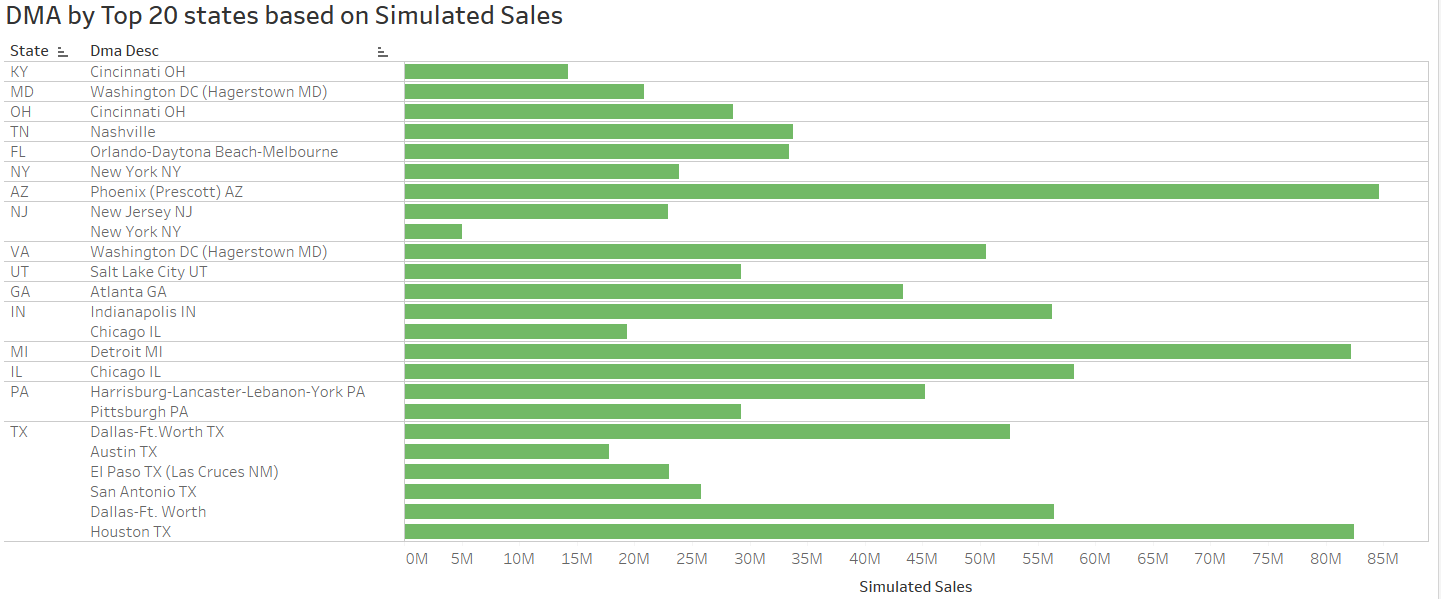
* Data Integrity, while uploading the data into the BI tool
* Tableau Desktop is recommended for sensitive data
* Checking if the BI tool has a secure interface

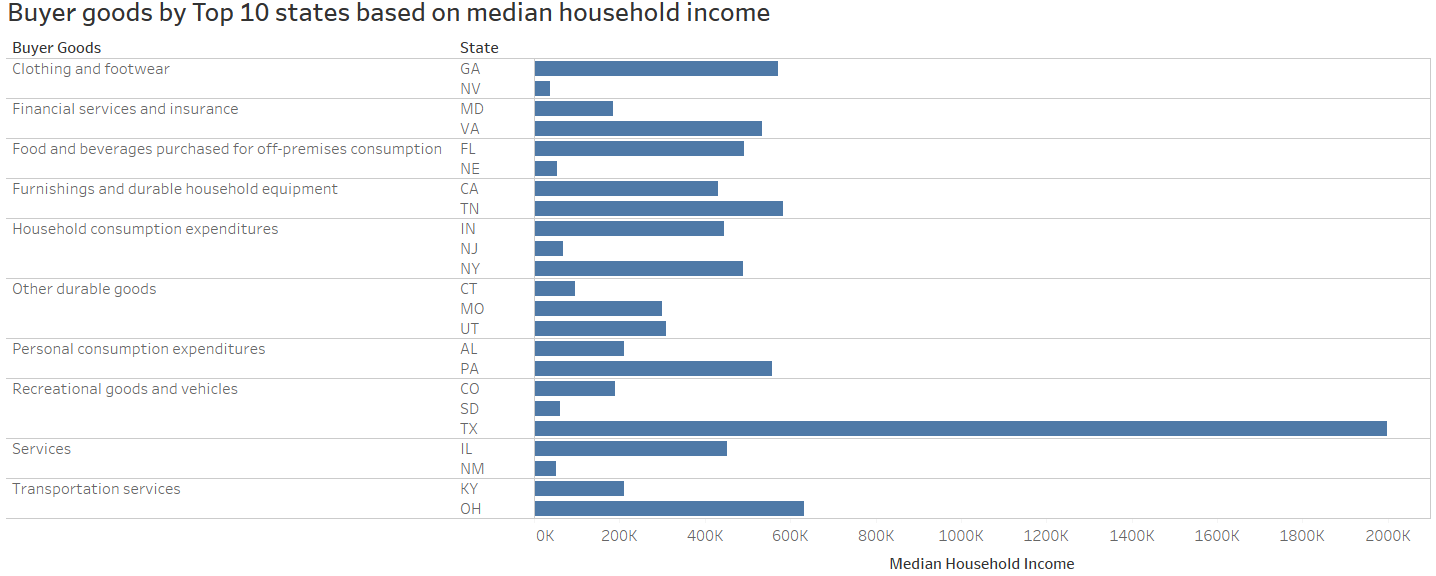


*Fig.1: Sample Dashboard showing the demographical information*

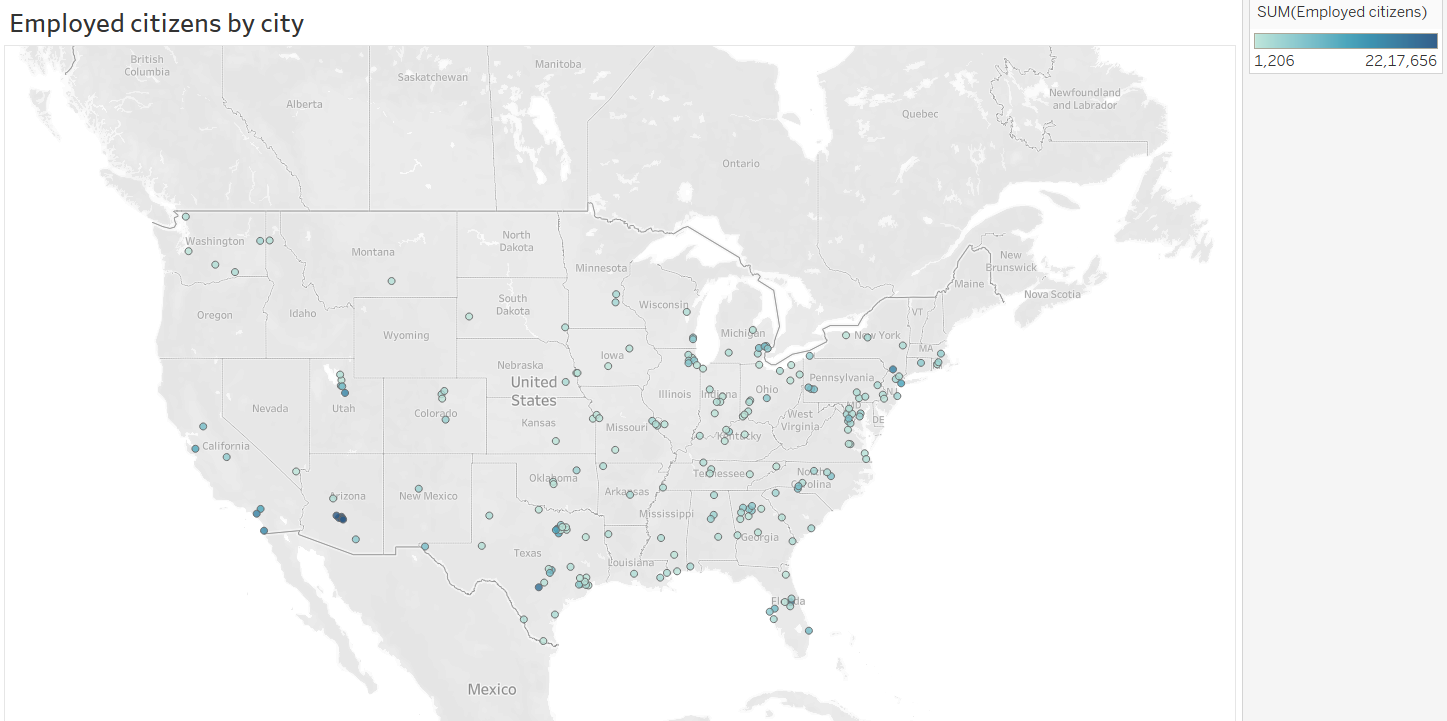
Statistical Tests:

* T-test: We can perform the t-test to measure the hypothesis of our samples. As we know, the t-test works by comparing the mean values of samples. Considering that, the population samples from our data have been tested for t-test by comparing their mean values. This analysis has helped in hypothesis testing to identify the sales in a given area.
* F-test: We have performed the f-test to compare the variances between the population samples. This can be performed by dividing the variance of means by the mean of group variance. This test helped in analyzing the hypothesis testing for the data, in identifying the revenue of the store over a given time, based on the customer sales in that location.
* Graphs:

*Fig.2: Designated Market Area by top 20 states based on the simulated sales*



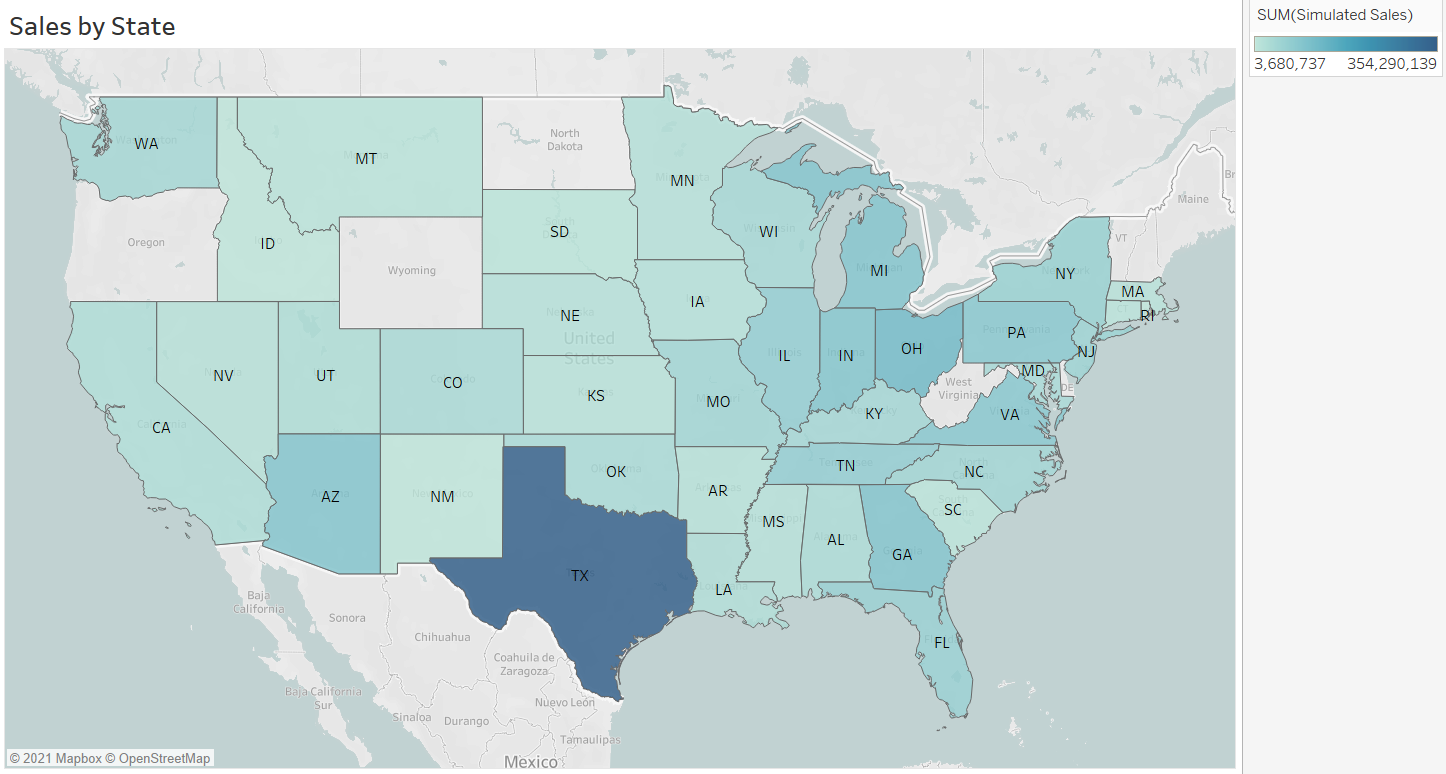
*Fig.3: Buyer goods by top 10 states based on median household income*



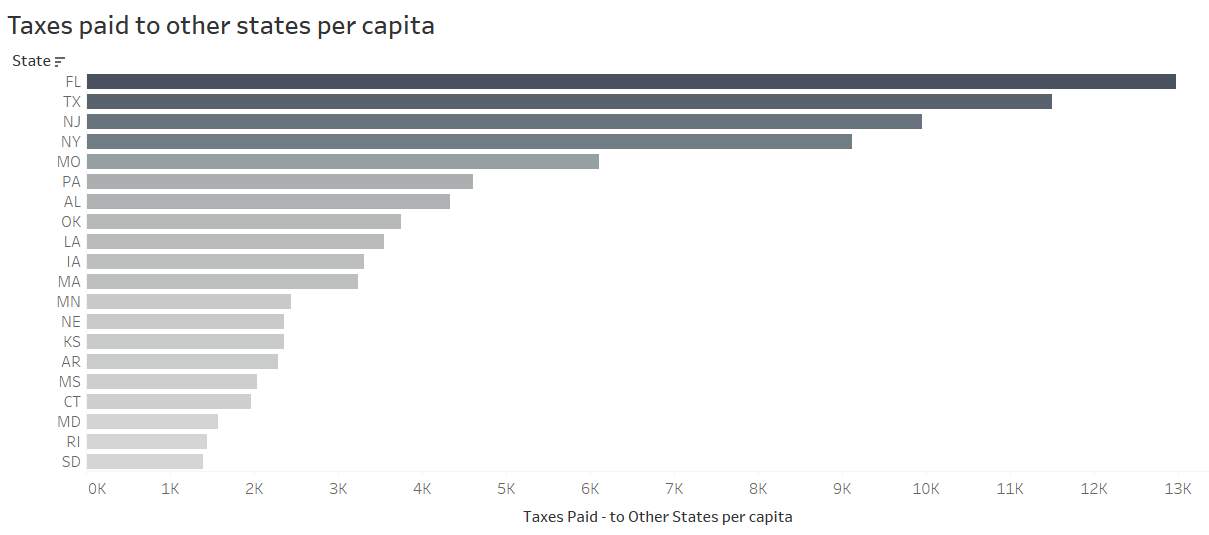
*Fig.4: Employed Citizens by City*



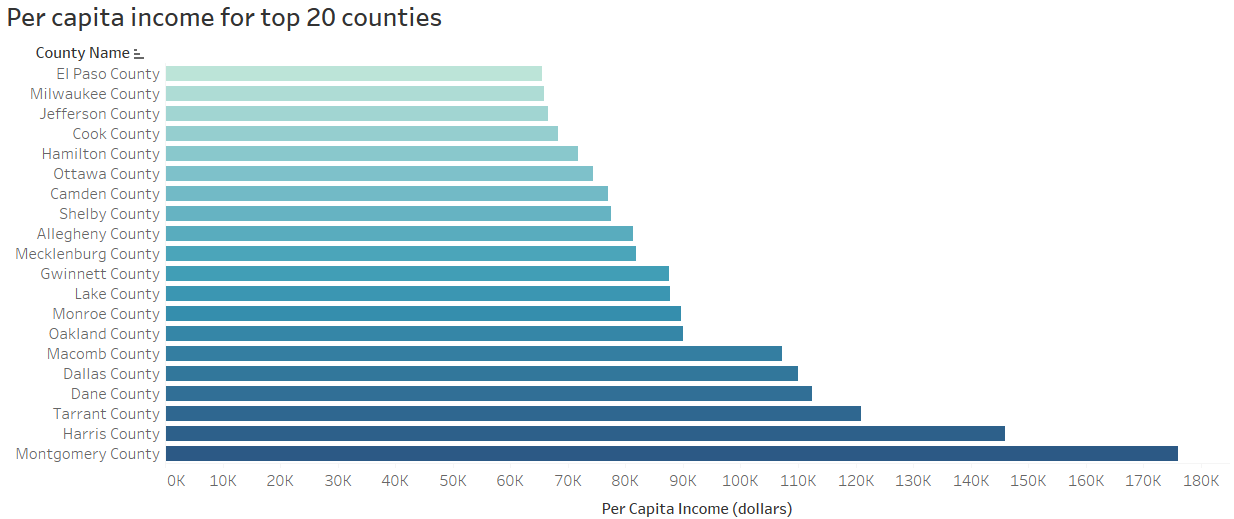
*Fig.5: Distribution showing the expenditure on consumer goods*



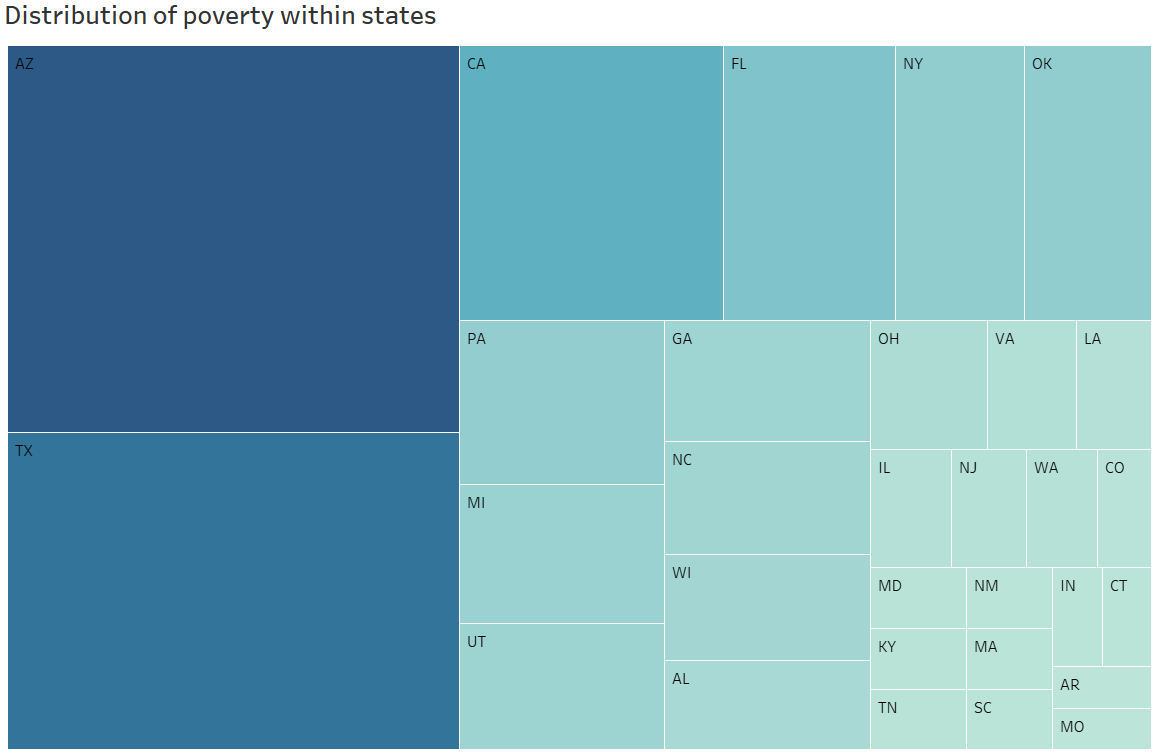
*Fig.6: Sales by State*



*Fig.7: Taxes paid to other states per capita*



*Fig.8: Per Capita Income for top 20 counties*



*Fig.9: Distribution of poverty within states*

Conclusion and Recommendations

After performing the project proposal, it can be concluded that the initial research and exploratory data analysis have helped in deciding the algorithmic approach for further analysis. Moreover, we have been considering the data governance perspective of implementation, which would be useful for implementing the intended data analysis. After analyzing the data visualizations, and results from statistical analysis, I intend to perform K-Means clustering on the data, as the data has population-based data, and cluster analysis will help in perfect model building for the data. Moreover, regression analysis would also help to understand the important parameters for our data. These implementations will be performed by considering the scope, data governance, and ethics for the At Home Group’s data. After analyzing the data visualizations, it was observed that:

* Arizona has the highest level of poverty, while Missouri has the lowest level of poverty
* The highest income has been noted in Montgomery County, while, lowest per capita income was noted in El Paso County.
* Phoenix in Arizona has observed the highest sales
* Texas has higher amount of sales

Business Recommendations

* Market Segmentation Analysis: It is important to understand the buying patterns of people to gauge the sales within a region. Therefore, market segmentation analysis will help to implement analysis based on customer buying patterns.
* Value Chain Analysis: This analysis is very important to improve profits using strategic methods. Most companies use this analysis by analyzing activities within the business. This helps in generating profits for the goods sold. Some companies perform this by selling the goods for a higher amount than their actual value.
* Vendor analysis: This analysis will help to improve sales and revenue, by employing the right suppliers for the business. This can be done by identifying suppliers that can yield the higher business for less amount of money.
* State-wise sales analysis: This analysis will help in identifying the sales patterns over time which will help in analyzing profits and losses made by stores.
* Analyzing the demographic patterns within the different areas: We can perform analysis on the employment rates, education rates, taxes per state, household income, etc., and compare with different states for performing state-wise sales analysis.
* E-commerce website: Enhancements in the e-commerce website and development of the mobile application will help in improving the sales for the store.

**References**

Zwitter, Anej J, & Hadfield, Amelia. (2014). Governing big data. Politics and Governance, 2(1), 1-3.

This article contains information about the importance of data governance in handling volumes of data. Moreover, the article talks about the techniques used to handle data governance and ethics while managing voluminous amounts of data.

Winter, Jenifer Sunrise, & Davidson, Elizabeth. (2019). Big data governance of personal health information and challenges to contextual integrity. The Information Society, 35(1), 36-51.

This article highlights the importance of compliance while managing healthcare data. Moreover, the article talks about the consequences caused due to misuse of security compliance while dealing with healthcare information.

Arthur, Keren Naa Abeka, & Owen, Richard. (2019). A Micro-ethnographic Study of Big Data-Based Innovation in the Financial Services Sector: Governance, Ethics and Organisational Practices. Journal of Business Ethics, 160(2), 363-375.

This article talks about the importance of data security and privacy while handling financial and sensitive information. Also, it focusses on the threats caused due to inconsistent data.