We would be looking for the students to identify any trends or insights from the customer sales, including seasonality of purchases or levels of purchasing throughout the year, the grocery bundles that are more often ordered together, or whether certain bundles are more commonly ordered in certain zip codes/areas. Overall, we are hoping to learn more about our customers' buying patterns.

**Project Timeline: September – November**

**Week 0:** ensuring data privacy and security are maintained

**Week 1-3: Project Introduction & Data Collection**

1. **Project Kickoff Meeting**: Introduce the project scope, objectives, and deliverables.
2. **Understanding the Business Context**: Discuss the client's business, the importance of customer sales analysis, and the expected outcomes.
3. **Initial Data Exploration**: Perform basic exploratory data analysis (EDA) to understand the structure, variables, and any obvious trends or issues.
4. **Data Cleaning**: Address missing values, outliers, and inconsistencies in the dataset.
5. **Data Transformation**: Prepare the data for analysis by creating new variables if needed (e.g., month, season, geographic regions).

**Week 4-6: Exploratory Data Analysis (EDA)**

**1. Data Overview and Initial Insights**

* **Understand the Dataset**:
  + Review the dataset to understand the variables included, such as product names, quantities sold, sale dates, customer IDs, and geographic information like zip codes.
* **Data Segmentation:** Segment the data based on key factors like time (months, quarters) and geography (zip codes, regions).
* **Descriptive Statistics**:
  + Calculate basic statistics (mean, median, mode, standard deviation) for key variables like sales volume, total revenue, and the number of transactions.
* **Data Distribution**:
  + Visualize the distribution of sales data using histograms, box plots, or density plots to identify patterns and outliers.
* **Time Series Analysis**:
  + Plot the sales data over time to identify any visible trends, seasonality, or cyclic patterns.

**2. Seasonality Analysis**

* **Monthly/Quarterly Sales Trends**:
  + Aggregate the sales data by month or quarter and create line charts or bar graphs to observe seasonal trends.
  + Identify peak sales periods and compare them with low sales periods to understand seasonality.
* **Year-over-Year Comparison**:
  + Compare sales data for the same periods across different years to identify consistent seasonal trends or anomalies.
* **Seasonal Decomposition**:
  + Use seasonal decomposition techniques (e.g., moving averages or time series decomposition) to separate the trend, seasonal, and residual components of the sales data.

**3. Purchase Patterns**

* **Product Sales Trends**:
  + Identify the top-selling products and analyze their sales trends over time.
* **Customer Segmentation**:
  + Segment customers based on purchase behavior (e.g., frequent buyers vs. occasional buyers) and analyze their contribution to overall sales.
* **Transaction Frequency**:
  + Analyze the frequency of purchases to identify customer retention and repeat purchase behavior.

**4. Bundling Analysis**

* **Association Rules**:
  + Use association rule mining (e.g., Apriori algorithm) to identify product combinations that are frequently bought together.
  + Generate rules like “If a customer buys Product A, they are likely to buy Product B” to understand common bundles.
* **Market Basket Analysis**:
  + Perform market basket analysis to identify product bundles that drive the most revenue.
  + Visualize common product bundles using heatmaps or network graphs.

**5. Geographic Analysis**

* **Sales by Region**:
  + Aggregate sales data by geographic regions (e.g., zip codes, cities) and create maps or choropleths to visualize sales distribution.
* **Hotspot Analysis**:
  + Identify areas with high sales activity (hotspots) and compare them with areas of low sales activity.
* **Geographic Segmentation**:
  + Segment customers based on their location and analyze the purchasing patterns in different regions.

**Week 7-8: Advanced Analysis & Insights**

**1. Correlation and Association Analysis**

* **Correlation Matrix**:
  + Create a correlation matrix to identify relationships between different variables, such as product sales, geographic locations, and time of purchase.
  + Visualize the correlations using heatmaps to identify strong relationships or dependencies.
* **Cross-Tabulation**:
  + Perform cross-tabulation analysis to explore the relationship between categorical variables, such as product categories and regions.

**2. Predictive Analysis**

* **Time Series Forecasting**:
  + Apply time series forecasting models (e.g., ARIMA, Exponential Smoothing) to predict future sales based on historical data.
  + Forecast future trends for top-selling products or specific regions to help the business plan inventory and marketing strategies.
* **Regression Analysis**:
  + Use regression models to predict sales based on factors like time of year, geographic location, and product type.
  + Identify the key drivers of sales and estimate their impact on future revenue.

**3. Customer Segmentation & Profiling**

* **Cluster Analysis**:
  + Use clustering techniques (e.g., K-Means, hierarchical clustering) to segment customers based on purchasing behavior, demographics, or geographic location.
  + Profile each segment to identify distinct characteristics, such as high-value customers, price-sensitive buyers, or seasonal shoppers.
* **Lifetime Value Analysis**:
  + Calculate customer lifetime value (CLV) for different segments to identify the most valuable customers and tailor marketing strategies accordingly.

**4. Insight Synthesis**

* **Key Findings Summary**:
  + Summarize the key insights from the analysis, focusing on actionable recommendations for the business.
* **Opportunity Identification**:
  + Identify opportunities for growth, such as new product bundles, targeted marketing campaigns, or regional expansion.
* **Risk Analysis**:
  + Highlight potential risks, such as declining sales in certain regions or products with decreasing demand, and suggest mitigation strategies.

**Week 8-9: Designing the Shiny App**

1. **Shiny App Introduction**: Introduce Shiny and its components, discussing the purpose and design of the app.
2. **App Layout Planning**: Plan the layout of the Shiny app, deciding on the key sections, visuals, and interactive elements.
3. **Basic App Structure**: Start building the basic structure of the app, including the user interface (UI) and server components.

**Week 10-11: Shiny App Development**

1. **Data Visualization**: Integrate the EDA results into the Shiny app using interactive charts, graphs, and maps.
2. **User Interactivity**: Add interactive features, such as filters and selectors, to allow users to explore the data based on their preferences.
3. **Testing and Debugging**: Test the app for functionality, usability, and performance, making necessary adjustments.

**Week 12: Finalization and Presentation**

1. **App Finalization**: Complete the final touches on the Shiny app, ensuring it's user-friendly and fully functional.
2. **Report Writing**: Prepare a final report summarizing the findings, insights, and recommendations based on the analysis.
3. **Client Presentation Preparation**: Develop a presentation to showcase the Shiny app and the key findings to the client.
4. **Final Presentation**: Present the Shiny app and project findings to the client and other stakeholders.

**Deliverables**

* **Shiny App**: An interactive web application displaying the results of the data analysis.
* **Final Report**: A comprehensive document summarizing the analysis, insights, and recommendations.
* **Presentation**: A slide deck used to present the findings to the client.

**Milestone Check-ins**

* **Biweekly Progress Reviews**: Regular check-ins to ensure the project is on track and address any issues or challenges.