Melissa Young
DSC680 – T301 Applied Data Science
8.2 Project 2: White Paper Milestone 3

Business Problem: At BKs Financial Corporation, we face the challenge of operating multiple diverse business units each functioning with their own set of operational norms and market strategies. This existing fragmentation can lead to inefficiencies, duplicated efforts, and missed opportunities for leveraging cross-unit synergies. The primary objective of this project is to unearth and exploit these synergies, aiming to harmonize operations, amplify customer value, and drive innovative strategies across the corporation.

Background/History: The construction and home appliance industries have historically operated in closely linked ecosystems. Builders and appliance dealers often form partnerships to provide fully equipped new homes to buyers. Understanding the dynamics between these industries is crucial for optimizing sales strategies and improving market penetration.

Data Explanation: Data has been sourced from internal systems detailing operational metrics, sales and marketing data, customer demographics, and service portfolios over the past 3 years. Public financial records and customer feedback channels will supplement internal data.

1. Policy/Builder List from MAK Corp: This dataset includes detailed records of policies associated with various builders and projects managed by MAK Corp. It will be instrumental in identifying operational synergies between our construction financing services and builder projects. Sourced from internal system of record. https://github.com/Joven0218/DSC680---Applied-Data-Science/blob/main/Project%209%20/mak_table
2. Policy/Manufacturer List from CENT Corp: Comprising policy records linked to dealers and their sales data, this dataset from CENT Corp will help analyze potential overlaps and collaborative opportunities between our automotive financing division and vehicle dealerships. Sourced from internal system of record. https://github.com/Joven0218/DSC680---Applied-Data-Science/blob/main/Project%209%20/cent_table

Methods and Analysis:

Data Preparation and Analysis

The project commenced with the loading and initial inspection of two distinct datasets: df_mak and df_cent. The df_mak dataset contains detailed information about builders and construction projects, whereas df_cent includes data on appliance warranty sales from various dealers.

1. **Data Loading and Initial Exploration**:
    o **Builder Data (df_mak)**: Loaded from an Excel file, containing fields like CertificateNumber, WarrantyEnrollmentAppID, ActivatedAt, and Premium. Initial exploration involved understanding data types, non-null counts, and basic statistics of numerical fields.
    o **Appliance Warranty Data (df_cent)**: Also loaded from an Excel file, featuring details such as DealerName, StateCode, EffectiveDate, and ContractCostAmount. A similar exploratory data analysis was performed to assess data quality and structure.
2. **Data Cleaning**:
    o **Duplicate Removal**: Both datasets were checked for duplicate records using drop_duplicates(), ensuring that the analysis would be based on unique entries only.

   o **Data Type Standardization**: Ensured that state identifiers across both datasets (StateId in df_mak and StateCode in df_cent) were standardized by converting them to uppercase and stripping any extra whitespace.

3. **Data Merging and Aggregation**:
   o **State-Level Aggregation**: Premiums and contract costs were aggregated by state to analyze geographical distributions of market activity. This involved grouping by state and summing up premiums and contract costs, respectively.
   o **Data Integration**: The aggregated data from both datasets were merged on the state identifier to facilitate a combined analysis, allowing us to view the total premiums and contract costs side by side for each state.

Data Integration and State-Level Insights

The merged data provided a comprehensive view at the state level, showing total premiums collected by builders and total contract costs from appliance warranties. This integration allowed for a comparative analysis across states to identify where the highest business activities occurred and where potential market opportunities lay.

1. **Geographical Distribution of Market Activity**:
   o **High Activity States**: California and Florida emerged as high activity states with substantial total premiums and contract costs, indicating robust market engagement and the potential for targeted marketing and strategic business development.
   o **Lower Activity but High Potential States**: States like Texas and New York showed a discrepancy between premiums and contract costs, suggesting untapped potential or different market dynamics that could be leveraged through customized strategies.
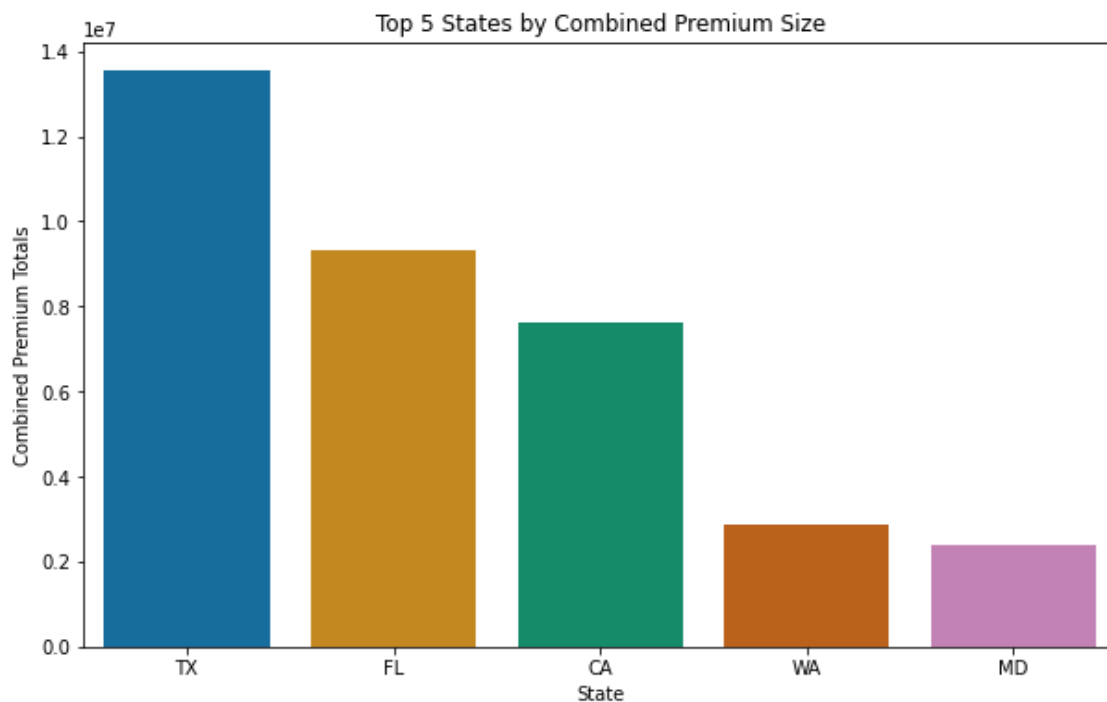
Statistical Analysis and Visualization

Using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, the project conducted various statistical analyses and visualizations:

1. **Descriptive Statistics**: Computed to get insights into central tendencies, dispersion, and shape of dataset distributions.
2. **Data Visualization**:
   o **Bar Charts**: Used to compare total premiums and contract costs across states.
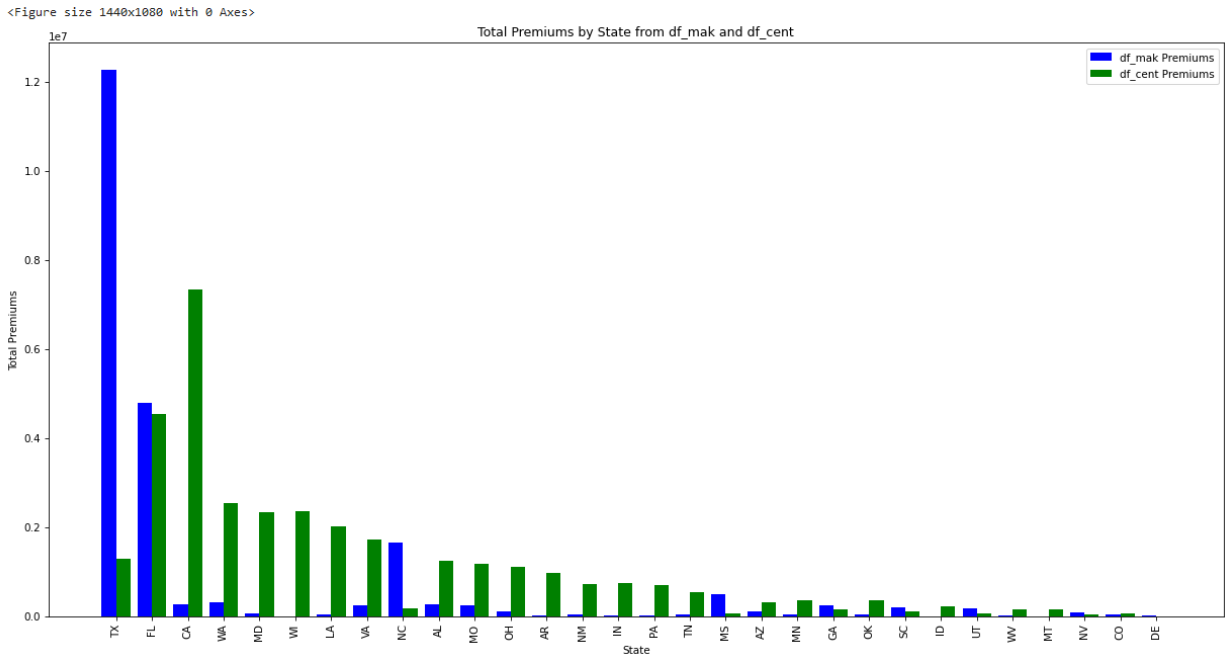

   o **Cluster Analysis Visualization**: Employed to identify patterns and clusters in the data based on premiums and costs.
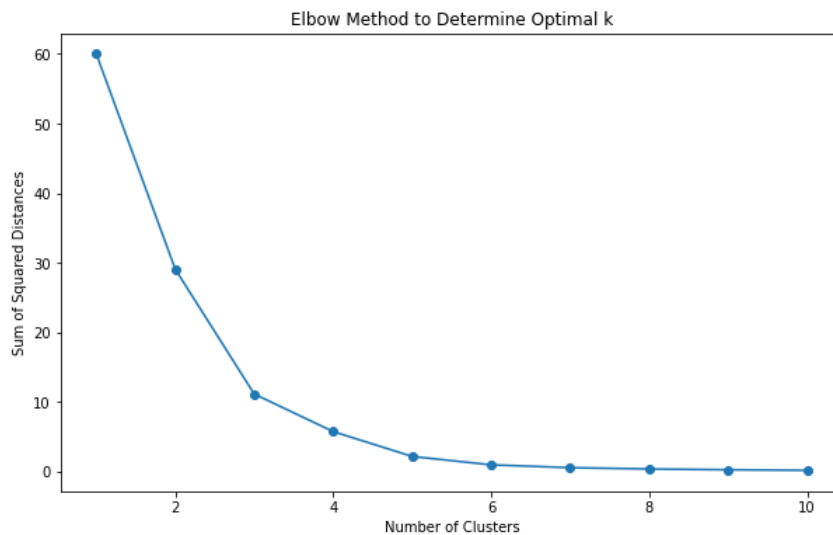
Predictive Modeling and Clustering

To further my analysis and predict future trends based on the gathered data, I applied the following methods:

1. **Standardization of Data**: Used StandardScaler from the sklearn library to normalize the data features, ensuring that the scale of the premiums and costs did not bias the clustering algorithm.
2. **K-Means Clustering**:
   o **Optimal Cluster Identification**: Applied the elbow method to determine the optimal number of clusters by analyzing the sum of squared distances for different values of k.
   o **Clustering Implementation**: Executed K-means clustering to segment states into clusters based on their premiums and contract costs dynamics.
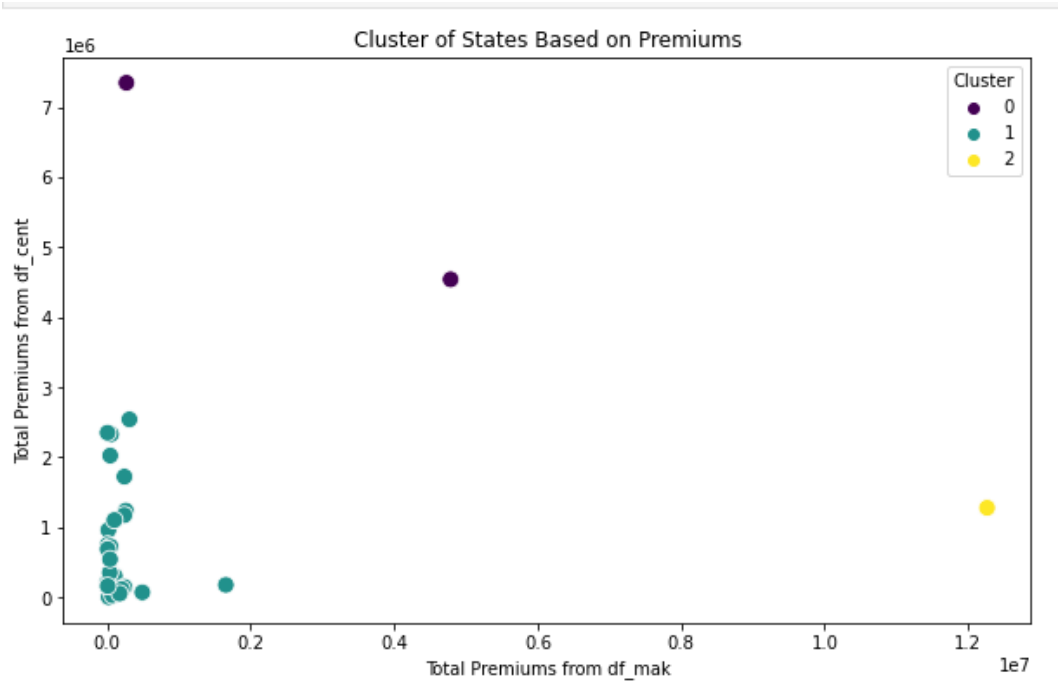
	Cluster	Total_Premiums_mak				Total_Premiums_cent \	
		mean	median	sum	count	mean
0	0	2.526142e+06	2526141.5	5052283	2	5.942943e+06
1	1	1.633413e+05	43754.0	4410216	27	7.548637e+05
2	2	1.226540e+07	12265403.0	12265403	1	1.278610e+06

	median	sum
0	5942943.48	11885886.96
1	360395.29	20381321.20
2	1278610.47	1278610.47

Cluster Analysis Summary:

Cluster 0

States: 2

Mean Premiums (mak): Approximately $2.53 million
Median Premiums (mak): Approximately $2.53 million
Total Premiums (mak): Approximately $5.05 million
Mean Premiums (cent): Approximately $5.94 million
Median Premiums (cent): Approximately $5.94 million
Total Premiums (cent): Approximately $11.89 million

Insights: Cluster 0 features states with very high average and median premiums from both datasets. The values are comparatively large, indicating that these states could be major markets with significant insurance activity or premiums volume.

Cluster 1

States: 27
Mean Premiums (mak): Approximately $163,341
Median Premiums (mak): $43,754
Total Premiums (mak): Approximately $4.41 million
Mean Premiums (cent): Approximately $754,863
Median Premiums (cent): $360,395
Total Premiums (cent): Approximately $20.38 million

Insights: This is the largest cluster by the number of states, featuring relatively lower mean and median premiums compared to Cluster 0. The premiums from df_cent are notably higher on average than those from df_mak, suggesting these might include states with lower direct premiums but higher associated costs or claims.

Cluster 2

States: 1
Mean Premiums (mak): Approximately $12.27 million
Median Premiums (mak): $12.27 million
Total Premiums (mak): $12.27 million
Mean Premiums (cent): Approximately $1.28 million
Median Premiums (cent): $1.28 million
Total Premiums (cent): $1.28 million

Insights: This cluster includes only one state but shows a very high total in df_mak premiums, massively outstripping the premiums recorded in df_cent. This suggests a unique condition where one type of premium or market activity is very dominant.

3. **Seasonality Analysis**:
   o **Time Series Extraction**: Converted dates to datetime format and extracted month and year for seasonal trend analysis.

    o **Aggregate Trends**: Analyzed monthly premiums and contract costs to identify seasonal patterns and variances over time.

Total Premiums and Contract Costs Combined by Month
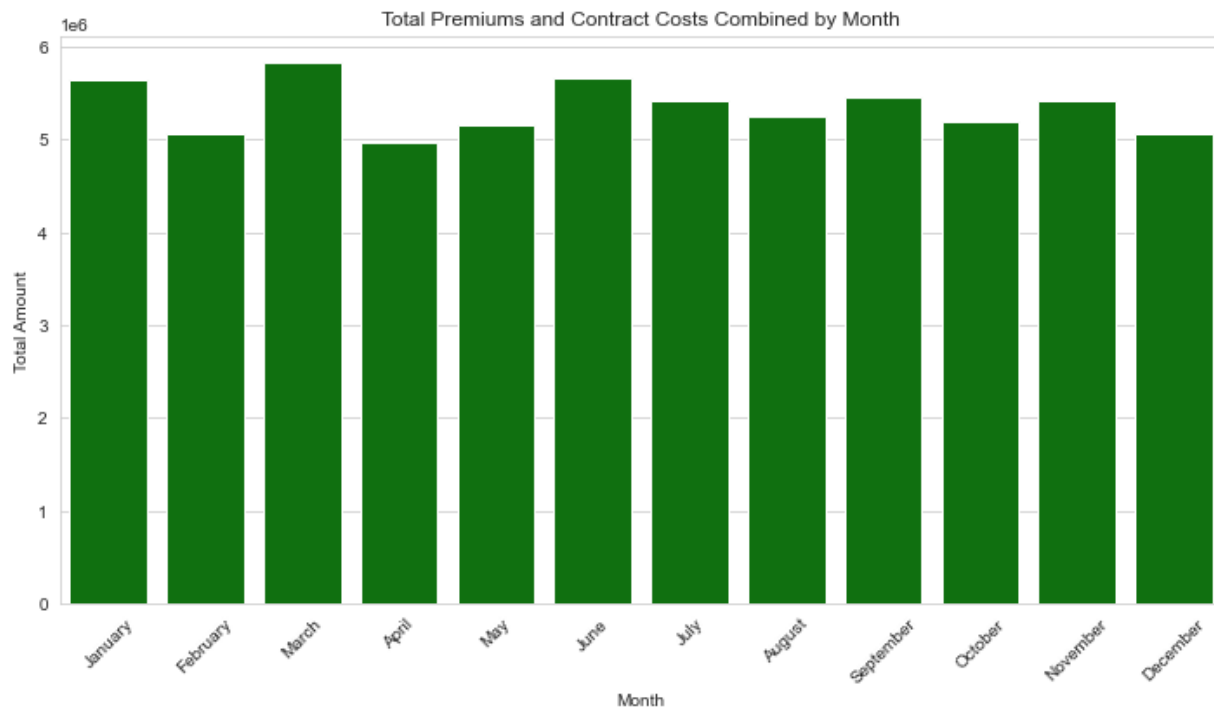
The seasonality analysis provided insights into how premiums and contract costs fluctuated throughout the year, which could assist in planning marketing strategies and inventory management.

1. **Key Findings**:
   - **Peak Months**: Both premiums and contract costs peaked during the spring and summer months, aligning with increased building activities and consumer purchases.
   - **Off-Peak Opportunities**: Identifying quieter months for targeted promotions could help in smoothing revenue streams and improving year-round sales consistency.

Predictive Analytics

The linear regression model aimed to predict future contract costs based on premiums data. The model provided insights but also highlighted the complexity of the market dynamics.

1. **Model Performance**:
   - **RMSE and R^2**: The model had a relatively high RMSE and low R^2, indicating that while there is a correlation between construction activities and appliance sales, numerous other factors influence the warranty sales which are not captured by premiums alone.

RMSE: 192693.67075581572

R^2: 0.18984687848028303

Actual vs. Predicted Contract Costs
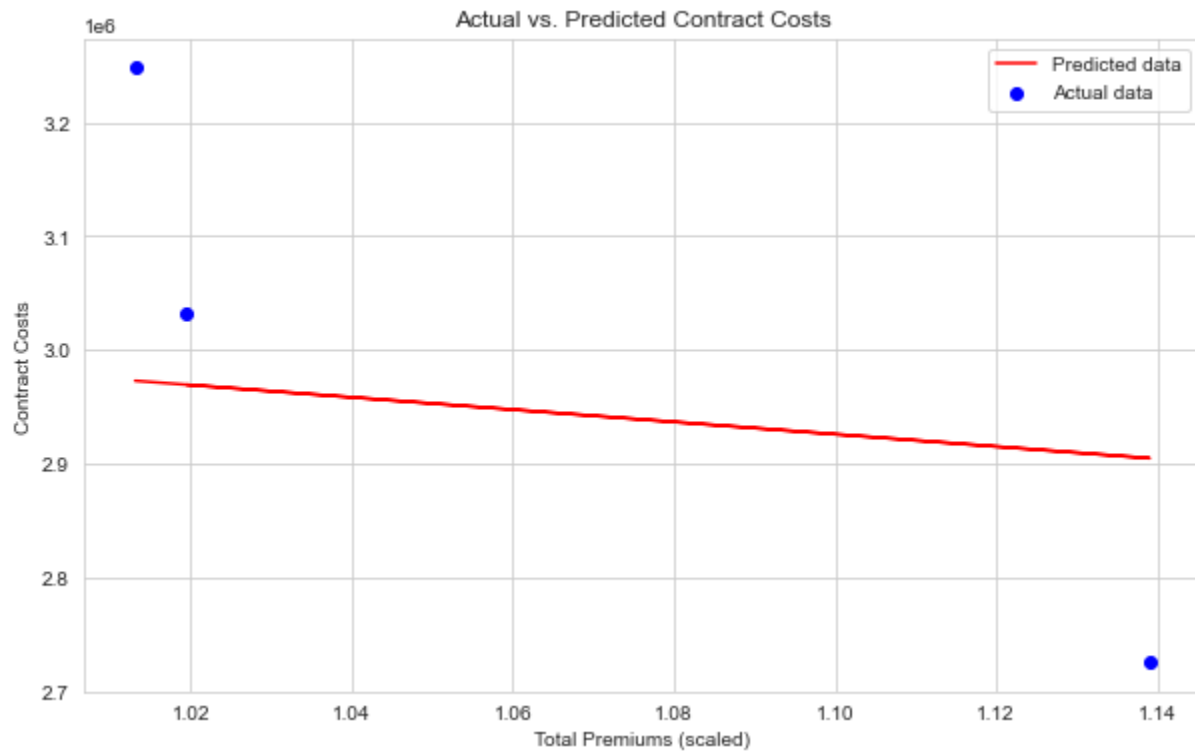
Conclusion: The analyses have demonstrated that strategic integration and targeted approaches tailored to specific states and market conditions can substantially improve both market reach and revenue streams. The identified high-activity states such as California and Florida, as well as the unique market dynamics in states like Texas, provide fertile ground for deploying focused marketing strategies and expanding service offerings.

As I move forward, it is imperative for BKs Financial Corporation to harness these insights to refine its strategic initiatives. The recommendations provided should serve as a blueprint for initiating change across units, encouraging collaboration, and innovating responses to market demands. By implementing these strategies, the corporation will not only achieve greater market penetration but also ensure sustainable growth and competitive resilience in a rapidly evolving industry.

Assumptions: The analysis of the datasets (df_mak for construction projects and df_cent for appliance warranties) relies on several key assumptions essential for interpreting the results and guiding decision-making. First, I assume that both datasets comprehensively capture all relevant transactions and activities within the specified analysis period, ensuring that all significant events or transactions that could influence the outcomes are included and accurately represented. Second, data collected across various business units and diverse geographic regions are presumed to maintain a consistent standard of quality and accuracy, which is crucial for valid comparisons and aggregations. Additionally, the insights and patterns uncovered are assumed to be generalizable beyond the specific datasets and timeframe examined, applying to similar periods, markets, or segments not directly scrutinized. I operated under the premise that observed correlations between construction activities (recorded in df_mak) and warranty sales (documented in df_cent) suggest causality, though it is recognized that correlation does not inherently establish causation. Lastly, the market conditions during the analysis period are presumed to be stable and representative of typical operational circumstances, ensuring that the predictive validity of the findings holds true under similar conditions without external anomalies or atypical events. These assumptions provide a clear

framework for understanding the boundaries and applicability of my findings, ensuring that the strategic recommendations derived are both robust and actionable.

Limitations:

1. **Data Scope and Availability**: One significant limitation might be the availability and scope of the datasets used (df_mak and df_cent). If the datasets do not capture all relevant variables or are limited to a specific geographic region or time period, the results may not fully encompass the complexity of the market dynamics.
2. **Historical Data Application**: The use of historical data to predict future trends inherently assumes that past patterns will continue into the future. This can be problematic in volatile markets or industries undergoing rapid change, where past performance may not be indicative of future results.
3. **Generalizability**: While the study assumes that insights and patterns can be generalized, this may not always hold true across different market segments or external conditions. Variations in economic, regulatory, or competitive landscapes could significantly influence the applicability of the findings.
4. **Data Quality and Integration**: Differences in how data is collected, processed, and stored across different business units can lead to inconsistencies that affect the integration and overall analysis. This could potentially skew results or obscure underlying trends.

Challenges:

1. **Complexity of Causal Inferences**: Establishing causality between construction activities and warranty sales is challenging without controlled experiments or more sophisticated statistical techniques that account for confounding variables. The correlation observed might be influenced by other unmeasured factors.
2. **Changing Market Conditions**: The assumption of stable market conditions is a significant challenge, particularly in industries like construction and retail, which are sensitive to economic shifts. External shocks such as financial crises, regulatory changes, or technological disruptions could invalidate the analysis.
3. **Resource Constraints**: Conducting a thorough and comprehensive analysis often requires substantial resources, including skilled personnel, advanced analytical tools, and sufficient time. Resource constraints can limit the depth and breadth of analysis, affecting the robustness of conclusions.
4. **Implementation of Recommendations**: Translating analytical insights into actionable strategies and implementing them across diverse business units involves complex coordination and change management. There may be resistance to change or differing priorities within the organization that challenge the implementation of recommended strategies.

Future Uses/Additional Applications:

Strategic Decision-Making

- **Expansion Planning**: The insights from the analysis can be used to guide expansion strategies into new markets or segments. By understanding which geographic areas or market segments demonstrate the highest synergies, strategic decisions regarding where to focus resources can be more effectively made.

- **Product Development**: Data showing strong correlations between construction activities and warranty sales might suggest opportunities for developing new products or services tailored to specific construction trends or customer needs.

Marketing and Sales Initiatives

- **Targeted Marketing Campaigns**: Knowledge of where and when the highest synergies occur allows for more targeted and effective marketing campaigns, designed to reach potential customers at optimal times.
- **Personalized Customer Engagements**: Insights into specific customer behaviors and preferences can lead to more personalized engagement strategies, enhancing customer satisfaction and loyalty.

Recommendations: Based on this cluster analysis, you might consider the following strategic actions:

Cluster 0: Focus on retention and enhanced services as these are likely mature markets with significant investment and interest.

Cluster 1: Investigate the reasons behind lower premiums in df_mak and higher costs or claims in df_cent. There might be opportunities for improving efficiency or adjusting pricing models.

Cluster 2: Due to its unique profile, further detailed analysis is necessary to understand why such a disparity exists between the two premium types. This state could be a key driver of specific business lines or products.

What If Scenario:

We know from the 2 datasets that the state of Florida is the synergistic state according to amount of premiums sold for both companies.

```
Largest Builder in Florida based on Premiums:
Company Name    Holiday Builders, Inc.
Premium                      490456

Largest Dealer in Florida based on Contract Costs:
DealerName           GOOD DEALS APPLIANCES
ContractCostAmount           1857578.06

Average Premium: 540.6161080838993

Average Contract Price Per Appliance: 75.45823637541326
```

If we apply the assumption of 908 homes (built per year by Holiday Builders) at an average contract price of $75*3 appliances = $ 204,300 in net contract premium increase if a relationship can be made between the builder and the appliance sales within the homes built.

1. Targeted Marketing and Sales Efforts

Holiday Builders, Inc.: Develop customized marketing strategies or incentive plans aimed at increasing the volume of business with this builder. Consider tailoring products or services to better meet their needs or offering bulk discounts to secure more premium volume.

GOOD DEALS APPLIANCES: Explore partnerships or exclusive deals that might be attractive to this dealer, given their significant role in the market. This could involve specialized warranty plans or service agreements that cater specifically to their customer base.

2. Relationship Building

Strengthen Relationships: Engage directly with the decision-makers at Holiday Builders, Inc., and GOOD DEALS APPLIANCES to understand their needs, expectations, and how your offerings can better serve their requirements. Strengthening these relationships can lead to more favorable terms and increased business.

Customized Offers: Based on the insights gained from interactions, develop customized offers or services that align with their current demands and future needs.

3. Market Analysis

Competitor Analysis: Investigate what competitors might be offering these companies and devise strategies to offer superior value.

Market Share and Penetration: Assess the market share of these entities in their respective sectors and explore ways to enhance it through joint ventures or co-marketing initiatives.

4. Operational Adjustments

Logistics and Supply Chain: For GOOD DEALS APPLIANCES, examine the logistics and supply chain efficiency to ensure that service delivery and product availability align with their business scale and demands.

Feedback and Product Development: Utilize feedback from these major players to adjust product or service offerings, which could lead to better market fit and customer satisfaction.

5. Financial Planning

Investment and Budgeting: Allocate marketing and sales budgets effectively to focus more intensively on these high-value customers. Consider investments in dedicated support teams or account managers to handle their needs.

Risk Management: Understand the risks associated with relying significantly on major players and consider strategies to diversify your client base to manage these risks effectively.

Implementation Plan:

- **Stakeholder Engagement**:
    - Identify and engage key stakeholders from various departments.
    - Establish clear communication channels for ongoing dialogue and feedback.
- **Resource Allocation**:
    - Allocate budget for necessary resources and technologies.
    - Assign dedicated personnel to oversee the implementation process.
- **Timeline Development**:
    - Develop a comprehensive timeline with specific milestones and deadlines.
    - Schedule regular review meetings to assess progress and make necessary adjustments.
- **Technology and Infrastructure**:
    - Select and acquire technology solutions that support data analytics and operational needs.
    - Ensure integration of new tools with existing systems for seamless data flow.
- **Training and Development**:
    - Provide training for team members on new tools and methodologies.
    - Implement a continuous learning program to keep skills and knowledge up-to-date.
- **Pilot Testing**:
    - Conduct a small-scale pilot to test strategies in a controlled setting.
    - Gather and analyze feedback to refine strategies before wider implementation.
- **Full-scale Implementation**:
    - Roll out the strategies gradually across the organization.
    - Monitor the implementation closely and adjust the strategy based on performance data.
- **Evaluation and Reporting**:
    - Define key performance indicators to measure the success of the implemented strategies.
    - Regularly report on performance to stakeholders and use findings for further refinement.
- **Feedback Integration**:
    - Integrate feedback from all stakeholders to enhance and optimize strategies.
    - Assess the scalability of successful strategies for broader application.

Ethical Assessment: In my exploration of synergies between construction projects and appliance warranty sales, it is paramount to uphold the highest ethical standards in data handling, analysis, and the application of insights. I am committed to ensuring data privacy and security, adhering to regulations such as GDPR and CCPA by securing explicit consent and employing robust security measures to protect data integrity. Transparency and accountability are cornerstones of my approach, requiring clear disclosure about data practices and establishing strict accountability to prevent and address any misuse.

Recognizing the importance of fairness, actively seek to identify and mitigate biases that may arise during data analysis or model implementation, ensuring my methodologies are fair and equitable across all customer groups. I also am acutely aware of the broader impact our data-driven decisions have on stakeholders, striving to assess and integrate their feedback to minimize negative outcomes and maximize societal benefits.

My commitment extends to promoting sustainability and social responsibility. I leverage my insights to encourage sustainable business practices and community well-being, aligning my operations with broader environmental and social governance objectives. Regular ethical reviews and ongoing education about evolving ethical standards and regulations ensure that my practices remain at the forefront of ethical compliance and corporate responsibility. Through these rigorous ethical practices, I aim to not only advance the business objectives but also contribute positively to the community and environment.

References:

Investopedia. (n.d.). Synergy. August 09, 2023, from https://www.investopedia.com/terms/s/synergy.asp

Kao, Clement. Understanding Product Synergies. Product Manager HQ. September 27, 2021, from https://medium.com/product-manager-hq/product-synergies-c4b06e06a415

https://github.com/Joven0218/DSC680---Applied-Data-Science/blob/main/Project%209%20/mak_table

https://github.com/Joven0218/DSC680---Applied-Data-Science/blob/main/Project%209%20/cent_table