Performane Evaluation Framework

Case Study: Taylor Company

Data Analytics for Performance Management
Assignment 2
Course Co-ordinator: Sasan Mansouri

Elisabeta Veliaj - S5792010

March 7, 2025

Faculty of Economics and Business
University of Groningen

Contents

1 Executive Summary **3**

2 KPI Definitions and Analysis **4**
2.1 Key Performance Indicators (KPIs) . . . 4
2.2 Potential External Data Enhancements . . . 6
2.2.1 Weather Data . . . 7
2.2.2 COVID-19 Lockdown Data . . . 7
2.2.3 Demographic Data . . . 7
2.2.4 Operational Data . . . 7
2.2.5 Geospatial Data . . . 8
2.3 Dataset Profiling . . . 8
2.4 Public Data Integration . . . 9
2.5 Exploratory Data Analysis . . . 10
2.5.1 Descriptive Statistics . . . 10
2.5.2 Breakdown Frequency Analysis . . . 10
2.5.3 Weather Condition Analysis . . . 11
2.5.4 Geographical Breakdown Analysis . . . 12
2.6 Distributor Performance Analysis . . . 12
2.6.1 External Factors on KPIs . . . 15

3 Bonus Allocation Framework **15**
3.1 Proposed Strategy . . . 15
3.2 External Factor Adjustments . . . 16
3.3 Final Performance Score Calculation . . . 17

4 Strategic Recommendations **18**
4.1 Operational Enhancements . . . 18
4.1.1 Route Optimization . . . 18
4.1.2 Workload Rebalancing . . . 18
4.1.3 Preventive Maintenance Scheduling . . . 19
4.2 External Data Utilization . . . 19
4.2.1 Weather Data Integration . . . 19
4.2.2 Demographic and Demand Data . . . 19
4.2.3 COVID-19 Lockdown Data . . . 19
4.3 Performance Communication . . . 20
4.3.1 Transparent KPI Reporting . . . 20
4.3.2 Performance Feedback Sessions . . . 20
4.3.3 Incentive and Bonus Communication . . . 20
4.3.4 Recognition Program . . . 20
4.4 Legal Developments . . . 20
4.4.1 Strategic Response . . . 21
4.4.2 Customer Retention Strategies . . . 21
4.5 Cultural and Regional Factors . . . 22
4.5.1 Proposed Methods for Adjusting Performance Evaluations . . . 22
4.5.2 Strategies for Region-Specific Support and Training . . . 23

5 Conclusion **23**

6 Appendix **25**
6.1 Detailed Results . . . 25
6.1.1 Table A: Performance Metrics per Distributor . . . 25
6.1.2 Table B: Breakdown Frequency by Weather Condition . . . 26
6.2 Plots . . . 27
6.2.1 Plot A: Breakdown Frequency Over Time . . . 27
6.2.2 Plot B: Breakdown Frequency by Weather Condition . . . 28
6.2.3 Plot C: Correlation between Weather Condition and Breakdown Frequency . . . 28
6.2.4 Plot D: Workload vs Efficiency . . . 29
6.2.5 Plot E: Efficiency vs Distance . . . 29
6.2.6 Plot F: Breakdown Frequency vs Distance . . . 30
6.2.7 Plot G: Breakdown Rate vs Distributor . . . 30
6.3 Code . . . 31
6.3.1 Code A: Python Code for Weather Data Integration . . . 31
6.3.2 Code B: Python Code for Analysis and Visuals . . . 33

1 Executive Summary

Taylor Company plays a pivotal role in maintaining and repairing McDonald's ice cream machines across the United States. The efficiency and reliability of Taylor's regional distributors directly influence the operational success of McDonald's franchises. Between October 2020 and December 2021, frequent machine breakdowns underscored the necessity for a comprehensive Performance Evaluation Framework. This report presents a detailed analysis aimed at enhancing distributor performance through the identification and evaluation of key performance indicators (KPIs), the development of a data-driven bonus allocation strategy, and the integration of external datasets to provide deeper insights.

Key Findings:

- **Response Time and Breakdown Rate:** Response time, defined as the duration between a reported breakdown and the initiation of repairs, is critical for minimizing downtime and maintaining customer satisfaction. Although precise response times were hindered by data limitations, proxies such as average distance to stores and workload ratios provided valuable insights. A positive correlation was identified between higher temperatures and increased breakdown frequencies, indicating that environmental factors significantly impact machine reliability.

- **Service Coverage and Proximity:** Distributors covering larger service areas tend to experience higher breakdown rates, suggesting that workload distribution plays a crucial role in performance. Efficient routing and proximity to stores emerged as significant factors in mitigating breakdown frequencies and enhancing service reliability.

- **Composite Efficiency KPI:** The composite efficiency KPI, which amalgamates response time, breakdown frequency, service coverage, and proximity, revealed substantial variability among distributors. High-performing distributors maintained manageable service areas and exhibited lower breakdown rates, highlighting the effectiveness of proactive maintenance and efficient resource allocation.

- **External Data Integration:** Incorporating weather data provided a nuanced understanding of how environmental conditions influence machine performance. The analysis demonstrated that severe weather conditions tend to reduce breakdown frequencies due to decreased machine usage, whereas moderate adverse weather correlates with higher breakdown rates owing to sustained demand.

- **Impact of Legal Developments:** Recent exemptions to the Digital Millennium Copyright Act (DMCA) have introduced competitive pressures by allowing third-party repair services. This shift necessitates strategic responses from Taylor Company to maintain its market position and revenue streams.

Bonus Allocation Framework: A structured bonus allocation strategy was proposed, weighting response time, breakdown frequency, and service coverage & proximity to compute a final performance score. This framework ensures fairness and incentivizes distributors to optimize their operations while accounting for external factors such as weather severity and lockdown impacts.

Strategic Recommendations:

- **Operational Enhancements:** Implementing advanced GIS and machine-learning-driven routing can optimize service routes, while workload rebalancing and preventive maintenance scheduling can reduce breakdown frequencies.

- **External Data Utilization:** Leveraging real-time weather forecasts and demographic data can enhance demand forecasting and resource allocation, ensuring distributors are well-prepared for demand fluctuations.

- **Performance Communication:** Transparent KPI reporting and interactive visualizations will foster better understanding and collaboration among distributors. Regular feedback sessions and recognition programs can further motivate distributors to achieve excellence.

- **Adaptation to Legal Changes:** Diversifying product offerings, enhancing customer support, and fostering partnerships with franchisees are essential strategies to counteract the competitive pressures introduced by the DMCA exemptions.
- **Cultural and Regional Adaptations:** Tailored training programs and region-specific support will address the unique operational challenges posed by varying cultural and regional practices, ensuring equitable performance evaluations.

In conclusion, by adopting a data-driven approach and implementing the recommended strategies, Taylor Company can significantly enhance the performance and reliability of its distributor network. This will not only improve operational efficiency for McDonald's franchises but also ensure sustained business growth and customer satisfaction in a dynamic market environment.

2 KPI Definitions and Analysis

To comprehensively assess the performance of Taylor Company's regional distributors, it is crucial to establish and analyze specific Key Performance Indicators (KPIs). This section delineates the selected KPIs, providing clear definitions, justifications, and the corresponding formulas. By evaluating these indicators, the aim is to identify areas of excellence and opportunities for improvement, thereby enabling data-driven decision-making and enhancing overall operational efficiency.

2.1 Key Performance Indicators (KPIs)

- **Response Time**
  - **Definition:** The time elapsed from when a machine breakdown is reported to when the repair process is initiated.

    Justification:
    * *Operational Efficiency:* Shorter response times indicate a distributor's ability to act promptly, minimizing downtime for stores and improving operational efficiency.
    * *Customer Satisfaction:* Faster initiation of repairs reduces the inconvenience faced by customers and staff, leading to higher satisfaction levels.
    * *Revenue Impact:* Ice cream products are a significant revenue stream. Reducing downtime ensures minimal revenue loss for McDonald's stores.
    * *Service Reliability:* Demonstrates dependable service from distributors.

    Formula:

    $$\text{Response Time (hours)} = \text{Repair Initiation Time} - \text{Breakdown Timestamp} \tag{1}$$

    Limitation:
    * *Data Availability:* Currently, the dataset lacks specific timestamps for when repairs are initiated and completed. Without this temporal data, calculating the exact response time is not feasible.
    * *Next Steps:*
      a) **Data Collection:** Implement a system to log precise timestamps for both breakdown reports and repair initiations.
      b) **Data Integration:** Ensure that the collected timestamps are integrated into the existing dataset for comprehensive analysis.
      c) **Process Automation:** Utilize automated tracking tools to minimize manual entry errors and enhance data accuracy.

- **Breakdown Rate (Reliability)**
  - **Definition:** The frequency of machine breakdowns at stores under each distributor's service area.

    Justification:
    * *Reliability Measure:* A lower breakdown rate suggests better maintenance practices and a focus on preventive measures by the distributor.
    * *Cost Control:* Frequent breakdowns increase repair costs and reduce machine lifespan. A lower rate translates to cost savings for both McDonald's and Taylor Company.
    * *Strategic Insight:* Identifying patterns in breakdown rates can help uncover systemic issues, such as specific machine vulnerabilities or regional usage differences.
    * *Operational Efficiency:* Fewer breakdowns reduce disruptions and repair costs.
    * *Customer Satisfaction:* Reliable machines ensure uninterrupted service, leading to higher satisfaction.
    * *Service Reliability:* Indicates consistent and high-quality maintenance services.

    Formula:
    $$\text{Breakdown Rate} = \frac{\text{Total Breakdowns}}{\text{Total Time Period (e.g., months or weeks)}} \tag{2}$$

- **Service Proximity (Geographic Coverage)**
  - **Definition:** The average distance between a distributor's base and the stores they service.

    Justification:
    * *Efficiency Indicator:* Distributors located closer to stores can respond more quickly to breakdowns, enhancing service reliability.
    * *Resource Optimization:* Shorter travel distances reduce fuel costs, travel time, and carbon footprint.
    * *Fair Workload Distribution:* Evaluating proximity helps identify whether service areas are assigned equitably among distributors, ensuring no distributor is overburdened due to long travel distances.
    * *Operational Efficiency:* Reduces resource usage and travel time.
    * *Customer Satisfaction:* Enables quicker service response, minimizing wait times.
    * *Service Reliability:* Ensures distributors are readily available to handle breakdowns efficiently.

    Formula:
    $$\text{Average Distance (km)} = \frac{\text{Sum of Distances}}{\text{Number of Stores Serviced}} \tag{3}$$

- **Service Coverage (Distributor Capacity)**
  - **Definition:** The number of stores within a distributor's assigned area.

    Justification:
    * *Capacity Assessment:* This KPI reflects whether distributors have the resources to manage the volume of stores effectively.
    * *Workload Balance:* Analyzing service coverage ensures distributors are not overextended, which could compromise service quality.
    * *Scalability Planning:* Understanding coverage helps Taylor Company assess whether current distributor assignments align with future growth or regional demand changes.
    * *Operational Efficiency:* Optimizes resource allocation and avoids overextension.
    * *Customer Satisfaction:* Ensures stores receive consistent and timely service.
    * *Service Reliability:* Balanced coverage supports steady and reliable operations.

Formula:

$$\text{Service Coverage} = \text{Count of Stores Serviced by Distributor} \tag{4}$$

- **Simplified Efficiency KPI**
  - **Key Assumptions for Simplified Efficiency KPI:**
    a) **Proximity Affects Efficiency**
      * **Rationale:** A closer distance between a store and its assigned distributor is likely to result in faster service, assuming consistent distributor performance.
      * **Simplified KPI:** Use average distance (`distance_km`) to distributor as a proxy for response efficiency.
      * **Formula:**

$$\text{Efficiency Score} = \frac{1}{\text{Average Distance to Distributor}} \tag{5}$$

      Higher scores indicate better efficiency (closer proximity).

    b) **Breakdown Frequency Reflects Efficiency**
      * **Rationale:** A lower breakdown frequency can indicate proactive maintenance and efficient service since frequent breakdowns may imply delays in repairs or ineffective fixes.
      * **Simplified KPI:** Use breakdown frequency per store as a proxy for distributor efficiency.
      * **Formula:**

$$\text{Breakdown Frequency per Store} = \frac{\text{Total Number of Breakdowns}}{\text{Number of Stores}} \tag{6}$$

      Lower values indicate better efficiency.

    c) **Number of Stores per Distributor Affects Efficiency**
      * **Rationale:** Distributors managing more stores might experience slower response times due to higher workload, potentially affecting efficiency.
      * **Simplified KPI:** Use number of stores per distributor as a proxy for workload and, indirectly, efficiency.
      * **Formula:**

$$\text{Workload Ratio} = \frac{\text{Number of Stores}}{\text{Number of Distributors}} \tag{7}$$

      Lower values indicate better efficiency (lower workload per distributor).

  - **Composite Efficiency KPI**
    * **Definition:** Combines multiple simplified proxies to reflect various aspects of service efficiency.
    * **Formula:**

$$\text{Composite Efficiency KPI} = w_1 \cdot \frac{1}{\text{Average Distance}} + w_2 \cdot \frac{1}{\text{Breakdown Frequency}} + w_3 \cdot \frac{1}{\text{Workload Ratio}} \tag{8}$$

    Where:
      · w_1, w_2, w_3 are weights that can be adjusted based on the perceived importance of each factor.

2.2 Potential External Data Enhancements

Integrating external data sources into the Performance Evaluation Framework can significantly enhance the depth and accuracy of the analysis. These potential data enhancements provide contextual insights that complement internal metrics, enabling a more comprehensive assessment of distributor performance and operational efficiency. While not all enhancements will be implemented in the current analysis, they are considered for future integration based on their relevance and data availability.

2.2.1 Weather Data

Purpose: Examine how weather patterns affect machine breakdowns and ice cream demand.

- *What's Needed:*
  - Temperature, and weather condition at store locations.
  - Seasonal analysis for trends in demand and machine strain.
- *Potential Gaps Addressed:* Helps adjust breakdown rates and demand forecasts for varying weather conditions.
- *Data Sources:* NOAA National Centers for Environmental Information, Weather Underground Historical Data, Open-Meteo.

2.2.2 COVID-19 Lockdown Data

- *Purpose:* Understand the impact of regional restrictions on service demand and breakdown frequency.
- *What's Needed:*
  - Timelines and severity of lockdowns across service regions.
  - Store-level traffic and usage data during lockdown periods.
- *Potential Gaps Addressed:* Adjust performance metrics during pandemic-affected periods.
- *Data Sources:* Johns Hopkins University, Our World in Data.

2.2.3 Demographic Data

Purpose: Correlate local economic and population factors with machine usage and sales trends.

- *What's Needed:*
  - Income levels, population density, and household sizes by service area.
  - Consumer purchasing behavior data for ice cream-related products.
- *Potential Gaps Addressed:* Provides context for breakdown rates and service coverage.
- *Data Sources:* U.S. Census Bureau, Local statistical agencies.

2.2.4 Operational Data

- *Purpose:* Integrate internal repair completion timestamps to calculate precise response times.
- *What's Needed:*
  - Access to maintenance records showing both breakdown timestamps and repair completions.
- *Potential Gaps Addressed:* Completes the formula for Response Time.
- *Data Sources:* Internal Taylor Company databases.

2.2.5 Geospatial Data

Purpose: Refine Service Proximity analysis by using precise geolocation.

- *What's Needed:*
  - High-resolution latitude and longitude data for both distributors and stores.
- *Potential Gaps Addressed:* Improves the accuracy of average distance calculations.
- *Data Sources:* OpenStreetMap, Google Maps API.

2.3 Dataset Profiling

To ensure a comprehensive analysis of distributor performance, the available datasets were carefully examined to understand their structure, completeness, and potential limitations. Below is a brief profiling of the key datasets used in the analysis.

McDonalStoresUS

Structure:

- Contains 10 columns with details about store locations and their proximity to distributors.
- Key fields: `store_id`, `distance_km`, `matched_stockist_id`.
- No missing values in any column.

Observations:

- Sufficient data on store location and distributor proximity for calculating service proximity (average distance).
- `distance_km` ranges from 0.27 km to 249.96 km, indicating variability in proximity.

Limitations:

- Does not include operational details like sales, customer traffic, or machine utilization.
- Static `distance_km` values do not reflect dynamic travel times or traffic conditions.

McDonalBrokensUS

Structure:

- Contains 2 columns: `store_id` and `timestamp`.
- Total records: 7,036,208 (extensive breakdown history).
- No missing values in either column.

Observations:

- Provides sufficient granularity for breakdown frequency analysis and trends over time.
- Timestamps are critical for calculating response times (when paired with repair completion data).

Limitations:

- Does not include repair completion times, preventing direct calculation of response time.
- Lacks details about breakdown causes or severity, which could help refine maintenance strategies.

TaylorUS

Structure:

- Contains 11 columns with distributor information.
- Key fields: `Stockist ID`, `Name`, `Service Area`, and `coordinates`.
- Missing values in:
  - `Phone` (1 missing).
  - `Email` (14 missing).
  - `Website` (1 missing).
  - `Service Area` (2 missing).

Observations:

- Includes qualitative service area descriptions, which may need geospatial mapping for precise analysis.
- Provides contact and operational details for distributors.

Limitations:

- Missing values in communication fields (`Email`, `Phone`, `Website`) could hinder distributor outreach.
- `Service Area` is vague; would benefit from geospatial polygons or precise boundary definitions.

2.4 Public Data Integration

Implementation: To enhance the analysis of distributor performance, city-level weather data were integrated using the Open Meteo API. The integration process involved fetching historical daily weather data—including weather condition, maximum temperatures, and minimum temperatures—for each McDonald's store location across the United States. To manage API rate limits and ensure reliable data retrieval, a caching mechanism was employed and retry logic. The retrieved data was then processed and combined into a single dataset for further analysis.

1. **Setup:** A caching mechanism (`requests_cache`) and retry logic were implemented to handle API rate limits and ensure reliable data retrieval.
2. **Parameter Definition:** The weather data was requested for the period from October 1, 2020, to December 31, 2021. The daily parameters included:
   - `weather_code` (categorical weather conditions)
   - `temperature_2m_max` (daily maximum temperature)
   - `temperature_2m_min` (daily minimum temperature)
3. **Data Retrieval and Processing:** For each store's city latitude and longitude, weather data was fetched, processed into a structured format, and combined into a single dataset.
4. **Output:** The final weather dataset was saved as a CSV file for subsequent analysis.

Limitations and Potential Impacts: Despite the comprehensive approach, certain limitations impacted the completeness of the weather data:

- **Incomplete Data Retrieval:** Due to daily API call limits, computational expense, and the shape of the resulting dataset, it was not possible to retrieve weather data for all cities. As a result, for cities where weather data could not be retrieved, the dataset was merged using a left join, resulting in `NaN` values for missing weather metrics.

- **Impact on Analysis:**
  - *Reduced Coverage:* Missing weather data for some cities reduces the overall dataset coverage, which may affect the accuracy of aggregated weather-based KPIs.
  - *Bias in Results:* Since weather data is missing for a subset of cities, there may be a bias in results if the missing data corresponds to specific geographic regions or climatic conditions.
  - *Potential Underestimation:* KPIs involving weather, such as breakdown rates influenced by extreme temperatures, may be underestimated in areas with missing data.

The detailed Python code for this integration is provided in Appendix 6.3.1.

2.5 Exploratory Data Analysis

This section presents an exploratory analysis of the dataset, focusing on temperature trends, weather conditions, and their impact on breakdown frequency. The analysis involves the examination of descriptive statistics, correlation plots, and temporal patterns.

2.5.1 Descriptive Statistics

The dataset consists of 4,596,884 records, each containing detailed information about store locations, weather conditions, and breakdown reports. Key statistics for relevant numerical features are as follows:

- Mean temperature ranges from $-31.66°$C to 43.48°C, with an average of 14.23°C.
- Mean time between Breakdowns varies significantly, with a mean of approximately 72 hours and a maximum duration exceeding 418 days.
- The distance to the nearest Taylor distributor has a mean value of 80.27 km, with a standard deviation of 65.66 km.

Missing values were identified in several fields, notably in the email (26.3%), website (1.1%), and weather-related features (0.018%). These missing values were either filled or excluded as appropriate during the analysis.

2.5.2 Breakdown Frequency Analysis

An examination of breakdown frequency over time indicates significant variability, with notable peaks occurring during periods of high temperature (Figure 7 in the Appendix). Scatterplots and regression analysis reveal a positive correlation between mean temperature and breakdown count, with a correlation coefficient of 0.24 (see Figure 1). This suggests that higher temperatures tend to increase the likelihood of equipment breakdowns. One plausible explanation for this trend is that during hot weather and periods of increased temperature, the demand for ice cream rises substantially, leading to more frequent and intensive use of ice cream machines. Consequently, the higher operational load on the machines during these periods results in an increased rate of mechanical failures and breakdowns.

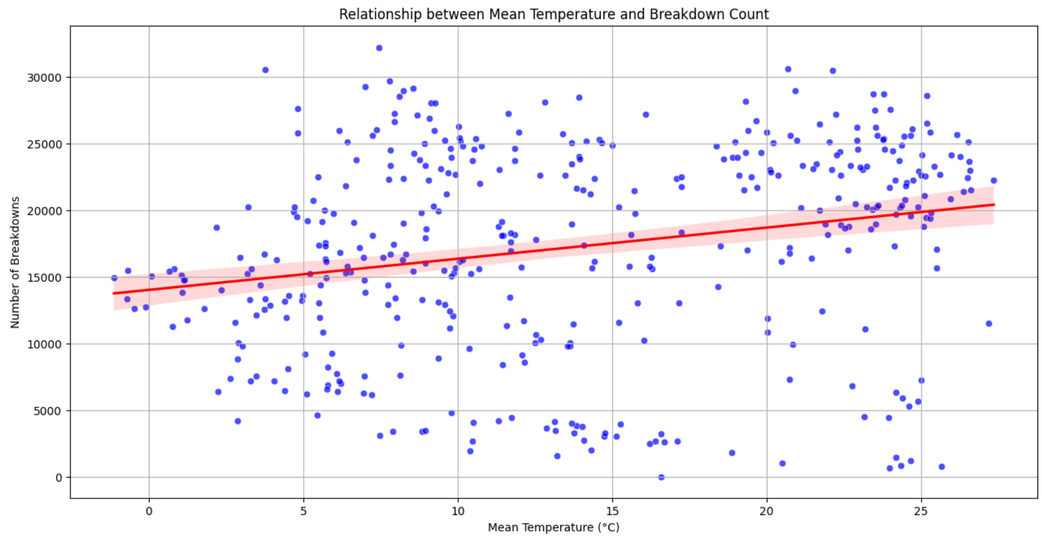

Figure 1: Relationship between Mean Temperature and Breakdown Count

2.5.3 Weather Condition Analysis

Weather conditions were categorized into **23 distinct types** based on weather codes, with **Sunny**, **Cloudy**, and **Foggy** being the most frequently observed conditions (see Table 2 and Figure 8 in the Appendix for a detailed breakdown). The analysis of breakdown frequency across these weather conditions shows a clear trend: breakdown frequency tends to decline slightly as weather conditions worsen. This trend is supported by a negative correlation coefficient of **-0.37**, indicating an inverse relationship between weather severity and breakdown frequency (see Figure 9 in the Appendix).

Interestingly, while severe weather conditions such as heavy rain and snow reduce breakdown frequency due to lower customer demand, some adverse weather types, like **Thunderstorms** and **Light Showers**, exhibit relatively high breakdown frequencies. This suggests that during moderate adverse weather, demand may remain high, resulting in continued machine usage. Conversely, during severe weather with reduced customer traffic, machines experience fewer operational cycles, lowering breakdown rates. Therefore, the negative correlation reflects the combined impact of weather-driven demand fluctuations and varying machine usage patterns.

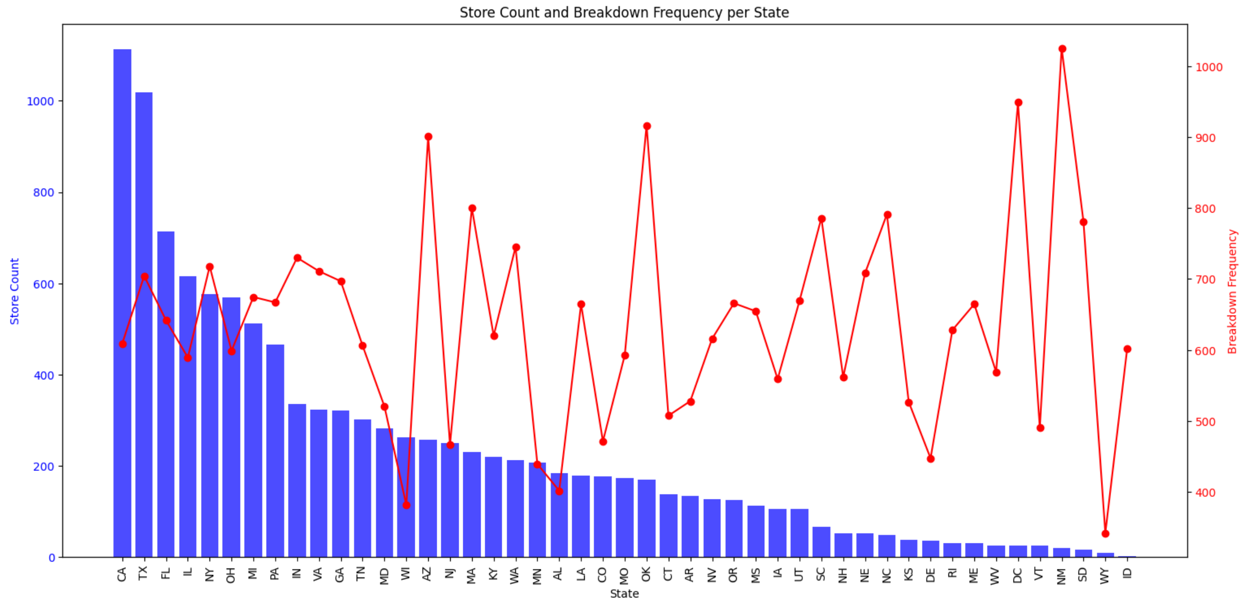

Figure 2: Store Count and Breakdown Frequency per State

2.5.4 Geographical Breakdown Analysis

Breakdown frequency was further analyzed across different states. California and Texas had the highest store counts and consequently reported the most breakdowns. However, states with fewer stores exhibited higher breakdown frequency rates. Figure 2: a dual-axis plot of store counts and breakdown frequency per state highlights this trend.

2.6 Distributor Performance Analysis

The Distributor Performance Analysis and the resulting plots reveal clear distinctions in how distributors manage their assigned territories:

- **Composite Efficiency KPI:**
  - Several distributors fall below a reference threshold of 0.30 (see Figure 3). These lower-scoring distributors tend to have large service coverage, high breakdown rates, or both. A threshold of mean was calculated.
  - High-efficiency performers (e.g., *Lane & McClain Distributors, Inc.* and *Taylor Freezer of Michigan, Inc.*) combine moderate-to-low breakdown frequency with manageable distances and workloads, indicating effective preventive maintenance or efficient service routing.

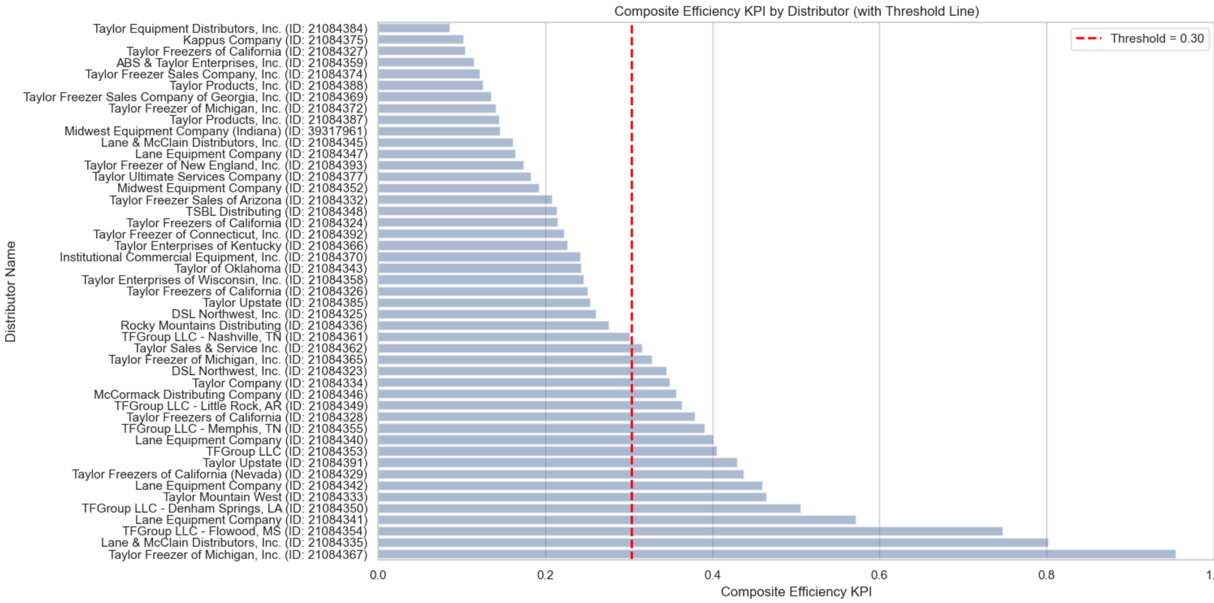

Figure 3: Efficiency by Distributor

- **Breakdown Rate vs. Service Coverage:**
  - As illustrated in the bubble plot (Figure 4), distributors with higher store counts show a corresponding increase in breakdown rate. This correlates strongly with larger workloads.
  - Figure 10 in the Appendix highlights how certain distributors servicing many stores in proportion to the total number of distributors (e.g., *Taylor Equipment Distributors, Inc.*) experience disproportionately high breakdown rates.

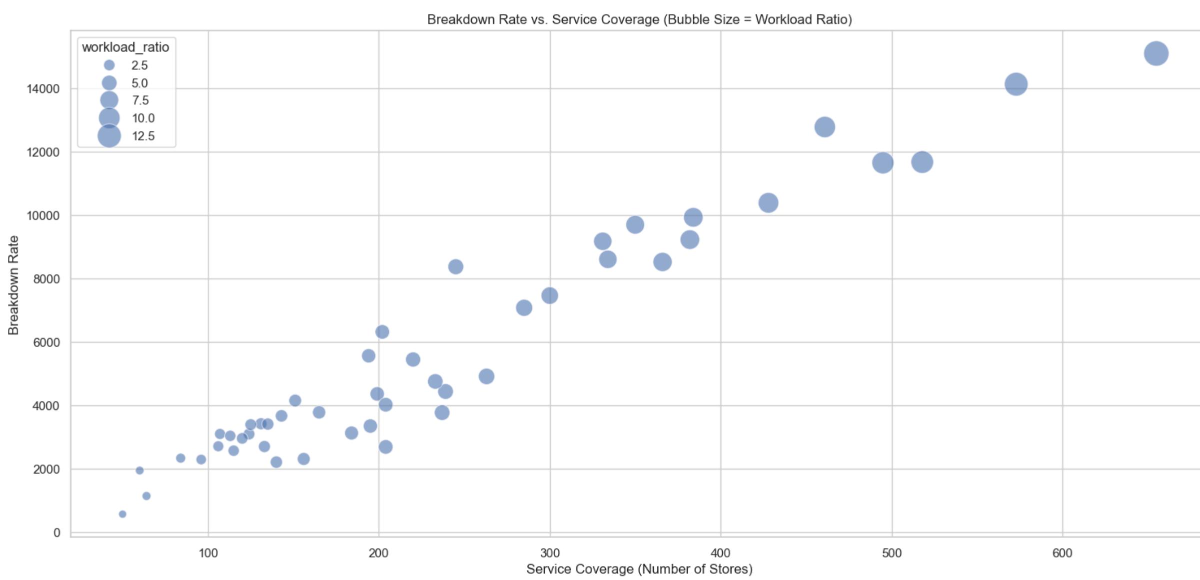

Figure 4: Breakdown Rate vs Service Coverage

- **Geographic Efficiency (Distance):**

- Based on the Table 1, in the Appendix which shows results per distributor, and Figures 11, 12 also in the Appendix, we conclude that while average distance influences breakdown frequency, the relationship is not absolute. Some distributors mitigate distance challenges likely via proactive maintenance or additional technician resources.
- Distributors like *Taylor Freezer of Michigan, Inc.* maintain high efficiency despite relatively higher distances, pointing to strong operational planning.

- **Correlations and Key Drivers:**
  - The pairwise plots (Figure 5) and heatmap (Figure 6) confirm service coverage and breakdown rate as the most tightly linked metrics, implying that large coverage alone often inflates total breakdowns.
  - Composite efficiency KPI is negatively correlated with volume-based factors (coverage, breakdowns), suggesting that expanding service areas without adequate support lowers overall efficiency.

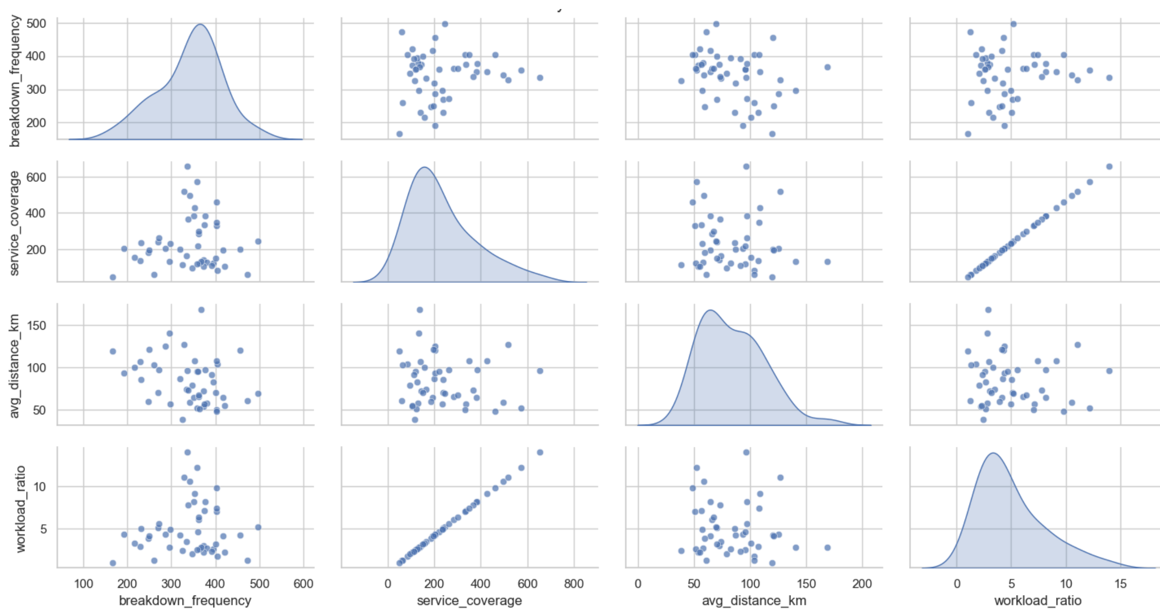

Figure 5: Pairplot

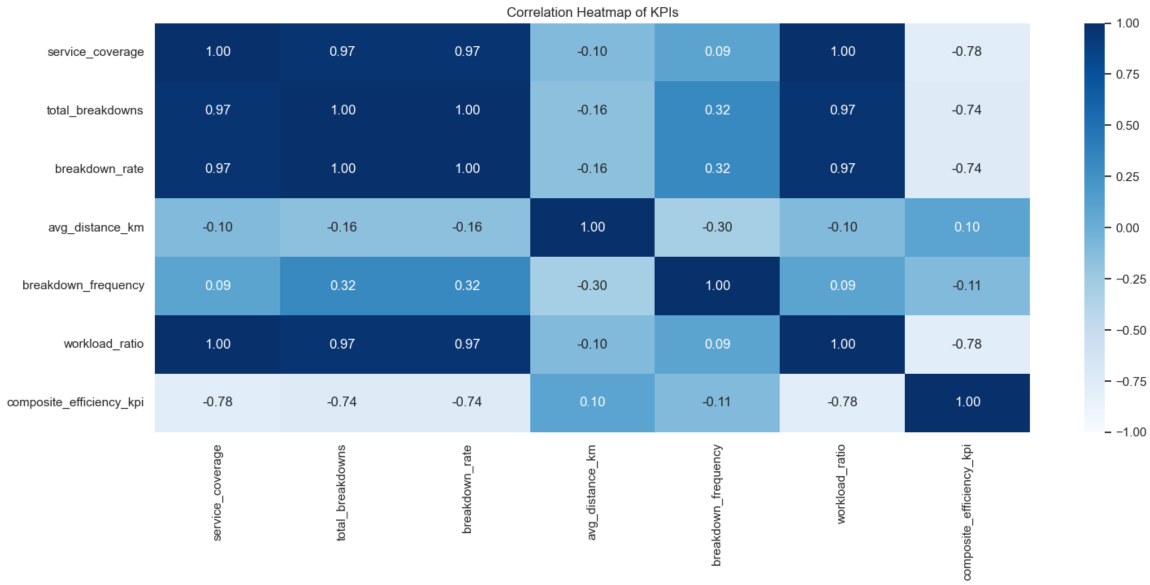

Figure 6: Correlation of KPIs

2.6.1 External Factors on KPIs

The analysis reveals that weather conditions significantly impact distributor performance, particularly through their effect on breakdown frequency. Higher temperatures correlate positively with breakdown frequency (0.24), likely due to increased machine usage during warm weather when ice cream demand peaks. Conversely, severe weather conditions, such as heavy rain or snow, correspond to lower breakdown frequency, as adverse weather reduces customer traffic and machine utilization.

In addition to weather, other external factors can significantly influence distributor KPIs. The COVID-19 pandemic and associated lockdowns likely caused fluctuations in machine usage, with reduced demand during lockdown periods and potential surges when restrictions eased. Similarly, distributors serving stores in touristic destinations may face seasonal spikes in demand, requiring flexible resource allocation. Demographics, such as local population density and income levels, can further drive ice cream sales and machine utilization patterns. Incorporating these factors into performance evaluation models can provide a more comprehensive view of distributor efficiency, helping tailor strategies to specific regional contexts.

3 Bonus Allocation Framework

3.1 Proposed Strategy

The proposed bonus allocation strategy aims to balance fairness, motivation, and adaptability by incorporating KPI scores for three primary metrics: **Response Time**, **Breakdown Frequency**, and **Service Coverage & Proximity**. Each metric is first converted into a *normalized score* on a scale where **higher = better**, then weighted to compute a final performance score.

- **Response Time (40%)**
  - *Definition:* The time between machine breakdown reporting and repair initiation.
  - *Normalization:* Because lower response time is better, define

$$\text{Response Time Score} = \frac{1}{\text{Response Time}} \times k_{\text{resp}},$$

where k_{resp} is a scaling constant (e.g., to bring typical response times into a 0–1 range).

– *Rationale:* Prompt response directly impacts downtime, store operations, and customer satisfaction. Although not calculated in the analysis due to data restrictions, response time remains a critical KPI for evaluating distributor performance. Accurate response time measurement requires access to detailed timestamps for both breakdown reports and repair initiation, which were unavailable in the current dataset. As a result, response efficiency was approximated using other metrics such as geographic proximity and workload. Future analyses should prioritize the collection and integration of repair timeline data to enable a more precise assessment of response time and its direct impact on service reliability and operational outcomes.

- **Breakdown Frequency Reduction (30%)**
  - *Definition:* Decrease in breakdown frequency per store over time (or a low absolute breakdown frequency).
  - *Normalization:* Since lower is better,

$$\text{Breakdown Frequency Score} = \frac{1}{\text{Breakdowns per Store}} \times k_{\text{break}},$$

    where k_{break} adjusts values to a comparable scale.
  - *Rationale:* Fewer breakdowns imply proactive maintenance, cost savings, and higher reliability.

- **Service Coverage & Proximity (30%)**
  - *Definition:* A combined measure of how many stores a distributor covers (capacity) and how close they are to each store (proximity).
  - *Sub-Weights:* 15% for Coverage, 15% for Proximity.

$$\text{Service Coverage Score} = \frac{\text{Coverage}}{\max(\text{Coverage})} \times k_{\text{cov}},$$

$$\text{Proximity Score} = \frac{1}{\text{Average Distance}} \times k_{\text{dist}}.$$

  - *Rationale:*
    * *Coverage:* Higher coverage may indicate strong capacity, but also must be supported by adequate resources.
    * *Proximity:* Being closer to stores reduces travel time and costs, improving response efficiency.

Weighted Sum: Once each raw metric is inverted or normalized to ensure "higher is better," the final score (prior to external adjustments) becomes:

$$\begin{aligned}\text{Performance Score}_{\text{raw}} = {} & 0.4 \times \text{Response Time Score} \\ & + 0.3 \times \text{Breakdown Frequency Score} \\ & + 0.15 \times \text{Service Coverage Score} \\ & + 0.15 \times \text{Proximity Score}.\end{aligned}$$

3.2 External Factor Adjustments

External factors such as severe weather, lockdowns, and demand variability can significantly impact performance. To ensure fairness, each of the main KPI scores is multiplied by a coefficient that *increases the final score* if the distributor has faced more difficult external conditions. Conversely, if conditions are more favorable, the coefficient is closer to or below 1. These adjustments ensure no distributor is unfairly penalized (or rewarded) for circumstances beyond their control.

1. **Weather Severity Adjustment** (C_{weather})

$$C_{\text{weather}} = 1 \;+\; \alpha_{\text{weather}} \times \left(\frac{\text{Weather Severity (Distributor Region)} - \text{Overall Avg. Severity}}{\text{Overall Avg. Severity}}\right).$$

   Interpretation:

   - If a region's weather severity is higher than average, $C_{\text{weather}} > 1$, boosting the *Breakdown Frequency Score* to offset uncontrollable weather impacts.
   - If the weather is milder, $C_{\text{weather}} \leq 1$, providing less of a boost.

   The factor α_{weather} controls the sensitivity of this adjustment.

2. **Lockdown Intensity Adjustment** (C_{lockdown})

$$C_{\text{lockdown}} = 1 \;+\; \alpha_{\text{lockdown}} \times \left(\frac{\text{Lockdown Days (Region)}}{\text{Total Days in Analysis Period}}\right).$$

   Interpretation: Distributors experiencing more lockdown days get a higher coefficient, ensuring they are not unduly penalized for slower response or fewer service calls when demand was restricted.

3. **Demand Variability Adjustment** (C_{demand})

$$C_{\text{demand}} = 1 \;+\; \alpha_{\text{demand}} \times \left(\frac{\text{Avg. Demand in Region} - \text{Overall Avg. Demand}}{\text{Overall Avg. Demand}}\right).$$

   Interpretation:

   - High-demand regions get a positive adjustment, recognizing the added stress on machines.
   - Low-demand regions receive a coefficient closer to or below 1.

Adjusted KPI Scores: Each KPI score is multiplied by the relevant external factor(s). For instance, if weather severely impacts breakdowns, we apply C_{weather} to the *Breakdown Frequency Score*:

$$\text{Breakdown Frequency Score}_{\text{adj}} \;=\; \text{Breakdown Frequency Score} \;\times\; C_{\text{weather}}.$$

If lockdown intensity mainly affects response times, we apply C_{lockdown} to the *Response Time Score*, etc.

3.3 Final Performance Score Calculation

After applying each relevant external factor, the final performance score is computed as:

$$\begin{aligned}\text{Performance Score}_{\text{final}} = {} & 0.4 \times \text{Response Time Score}_{\text{adj}} \\ & + 0.3 \times \text{Breakdown Frequency Score}_{\text{adj}} \\ & + 0.15 \times \text{Service Coverage Score}_{\text{adj}} \\ & + 0.15 \times \text{Proximity Score}_{\text{adj}}.\end{aligned}$$

Rationale for Weighting and Adjustments

- **Weights:**
  - *Response Time (40%)*: Highest weight due to direct effect on downtime and customer satisfaction.
  - *Breakdown Frequency (30%)*: Reflects maintenance quality and proactive service.
  - *Coverage (15%) & Proximity (15%)*: Ensures that total workload and geographic spread are accounted for.

- **Higher-is-Better Normalization:**
  - KPIs where lower raw values are better (e.g., Response Time, Breakdown Frequency) are *inverted*.
  - Coverage can be scaled directly if higher coverage is considered positive, or partially inverted if we want to avoid penalizing distributors covering large areas.
- **External Factor Adjustments:**
  - *Equity:* Distributors should not be unduly penalized for events beyond their control (e.g., harsh weather, strict lockdowns).
  - *Motivation:* Normalizing the performance metrics fosters fair comparison across diverse regions.

By formalizing the scoring methodology, normalizing the KPIs, and applying targeted external factor adjustments, this revised framework promotes both *fairness* and *operational excellence*. It ensures distributors remain motivated to improve the areas they can control while avoiding undue penalties for conditions outside their influence.

4 Strategic Recommendations

The following recommendations for strategic improvement synthesize the findings from the analysis done but also external factors influencing the company. While it may not be feasible to implement all recommendations simultaneously, adopting even a subset of these strategies will contribute to significant operational enhancements, foster equitable performance evaluation, and improve overall service reliability across Taylor Company's distributor network. By taking a phased approach, Taylor Company can prioritize key initiatives that yield the highest immediate impact while progressively building a more efficient and resilient service model.

4.1 Operational Enhancements

4.1.1 Route Optimization

As observed in Figure 11 and Table 1, some distributors maintain relatively high efficiency despite longer distances, suggesting that effective routing and proactive planning can mitigate the limitations of geography.

- **Leverage Advanced GIS and Traffic Analytics.** Use real-time traffic data, store operating hours, and historical breakdown frequency (see Section 2.6 for frequency trends) to determine the most efficient routes.
- **Implement Machine-Learning–Driven Routing.** Predictive models can forecast areas likely to experience higher demand (e.g., during hotter months, as shown by the temperature correlation in Figure 1) and prioritize these routes.
- **Incorporate Breakdown Severity and Response Windows.** Assign a priority score to each service call based on factors such as time-since-breakdown, store busyness, and local temperature spikes (refer to Section 2.2 for ideas on leveraging weather data).

4.1.2 Workload Rebalancing

Figure 10 illustrates how distributors servicing significantly more stores often face higher breakdown rates. The workload metric (Appendix Table 1) highlights which distributors may be overburdened.

- **Dynamic Workload Allocation.** Integrate automated alerts when a distributor's breakdown rate exceeds a threshold. Re-route or temporarily assign technicians from nearby, less-burdened distributors.

- **Seasonal and Event-Driven Reassignments.** Use historical data around high-temperature months (see correlation with breakdowns in Figure 1) or large-scale events to anticipate workload spikes and reassign resources as needed.
- **Capacity Planning Simulations.** Run "what-if" analyses (e.g., if demand rises 20%) to project how workloads could shift (Section 2.6 shows how coverage heavily influences breakdowns). Adjust service area boundaries preemptively for peak seasons.

4.1.3 Preventive Maintenance Scheduling

The correlation coefficient of 0.24 between higher temperatures and breakdown frequency (Figure 1) underscores the importance of proactive servicing during warmer periods.

- **Temperature-Triggered Maintenance.** Schedule extra maintenance during periods of high demand (summer months, heatwaves) by combining weather forecasts (Section 2.2) with historical breakdown patterns (Section 2.6).
- **Targeted Inspections After Peak Seasons.** Data in Figure 7 suggests breakdowns often surge following sustained heavy use. Plan post-peak system checks or part replacements.
- **Automated Maintenance Alerts.** Install IoT sensors to detect strain (e.g., unusual temperature or pressure readings) so distributors can intervene before a full breakdown. This aligns with the idea of improving data collection for exact response-time measurements (Section 2.2).

4.2 External Data Utilization

4.2.1 Weather Data Integration

As demonstrated in Figure 1 and Table 2, weather conditions correlate strongly with breakdown rates.

- **Real-Time and Forecast Alerts.** Fetch daily or weekly temperature forecasts to anticipate high-demand periods. If a region is expected to exceed a temperature threshold, allocate extra maintenance crews (see also Section 2.2).
- **Weather Severity Coefficients.** Adjust distributors' KPI scores (Section 2.6 on KPI weighting) based on regional weather severity so that those operating in consistently hot or extreme climates are not unfairly penalized.

4.2.2 Demographic and Demand Data

Demographics can influence machine usage patterns.

- **Local Demand Forecasting.** Combine census and local tourism data with breakdown logs. High-density or tourist-heavy areas may see elevated machine usage (refer to high breakdown frequencies in certain states shown in Figure 2).
- **Demand Variability Index.** Develop an index (similar to the "demand variability adjustment" from Section 2.2) to reflect seasonal or event-driven spikes, informing resource allocation and bonus adjustments.

4.2.3 COVID-19 Lockdown Data

- **Lockdown-Adjusted KPIs.** Integrate a "lockdown intensity factor" (as suggested in Section 2.6 regarding external factors) so that distributors in regions with extensive restrictions are not penalized for reduced service demand.
- **Post-Lockdown Surge Planning.** Historical data (Figures 7 and 1) show how demand can spike once restrictions lift. Use past trends to plan resources and maintain robust service levels.

4.3 Performance Communication

4.3.1 Transparent KPI Reporting

By referencing the composite efficiency KPI framework (Section 2.6) and the bonus allocation strategy, distributors can better understand their performance drivers.

- **Monthly "Adjusted Score" Dashboards.** Present both raw and adjusted KPIs so distributors understand precisely how their performance is calculated. Refer to Figure 3 for an example of how to visualize efficiency scores.
- **Interactive Visualizations.** Allow filtering by temperature, store coverage, or maintenance logs to demonstrate how these variables affect breakdown frequency (similar to Figure 5 showing KPI correlations).

4.3.2 Performance Feedback Sessions

Comparisons like those shown in Figures 4 and 3 can stimulate discussion on best practices.

- **Data-Focused Workshops.** Feature sessions where high-performing distributors (e.g., those with below-average breakdown rates or above-average composite efficiency KPI; see Table 1) share strategies.
- **Peer Benchmarking and Best Practices.** Encourage direct comparison among distributors with similar coverage footprints and discuss operational tactics that drive efficiency, referencing the threshold metrics shown in the pairplots (Figure 5).

4.3.3 Incentive and Bonus Communication

Section 2.6 and the *Bonus Allocation Framework* outline how KPI weights and external factor adjustments affect final scores.

- **Regularly Update External Factor Coefficients.** Since weather, lockdowns, and demand data can fluctuate, publish any coefficient updates or weighting changes at least quarterly (see Section 2.2 for data sources).
- **Illustrative Scenarios.** Provide examples of how extreme weather or lockdowns can alter a distributor's bonus calculation, improving transparency (Section 2.6 offers guidance on external factors).

4.3.4 Recognition Program

Recognition should be rooted in the objective measures explained in Section 2.6 (e.g., highest improvement in breakdown rate or best management of high-demand periods).

- **Highlight Seasonal Champions.** Introduce quarterly awards for distributors who successfully navigated extreme conditions (e.g., "Heat Wave Hero") consistent with the observed temperature-driven spikes (Figure 1).
- **Spotlight Operational Innovations.** Reward creative solutions such as IoT-based preventive maintenance (Section 2.2 suggests operational data improvements) or regional micro-depots that reduce travel time and breakdown severity.

4.4 Legal Developments

The recent U.S. Copyright Office ruling, which exempts retail-level commercial food equipment from the Digital Millennium Copyright Act (DMCA), is a major shift for Taylor Company. Part of the broader "Right to Repair" movement, this exemption allows businesses to repair their own equipment without relying solely on manufacturers. Previously, Section 1201 of the DMCA barred bypassing digital locks on software, granting Taylor a near-monopoly on McDonald's ice cream machine repairs and significant revenue from service contracts (Shepherd, 2024).

Taylor's C602 machines, known for frequent malfunctions, could only be repaired by Taylor-certified technicians due to restricted access to diagnostic tools. This approach forced franchisees to rely on costly Taylor services, increasing downtime and hurting sales during high-demand periods (Wade-Vicente, 2024).

The decision, driven by lobbying efforts from groups like iFixit, now enables franchisees to bypass digital locks and use third-party repair services, potentially disrupting Taylor's revenue model and increasing competition (Shepherd, 2024).

- **Impact on Distributors**
  The ruling significantly impacts Taylor's distributors, who previously benefitted from the company's monopolistic control over repairs. With the entry of independent repair providers and the potential adoption of alternative maintenance solutions by franchisees, distributors may experience reduced demand for their services. Estimates suggest that repair contracts contributed up to 25% of Taylor's annual revenue, indicating the magnitude of potential revenue loss (Wade-Vicente, 2024).

- **Competitive Landscape**
  The exemption opens the market to third-party repair providers, such as startups like Kytch, which had previously developed a device to monitor and optimize Taylor's machines. Although Kytch faced legal challenges and pushback from both Taylor and McDonald's, the recent ruling legitimizes the role of independent repair solutions (Hauptman, 2025). The growing presence of such competitors will force Taylor to innovate and potentially reduce repair costs to stay relevant.

4.4.1 Strategic Response

To navigate this shift, Taylor must diversify its offerings and improve service quality. Several strategic initiatives can help the company adapt:

- **Expanding Product Portfolio**: Taylor should invest in developing new, innovative food preparation equipment beyond ice cream machines, targeting broader markets.

- **Enhanced Customer Support**: By offering better service packages, including preventive maintenance and faster response times, Taylor can retain a competitive edge despite increased repair alternatives.

- **Collaboration with Franchisees**: Establishing co-creative partnerships with McDonald's franchise owners to develop user-friendly equipment and support systems may foster long-term loyalty.

4.4.2 Customer Retention Strategies

Retaining McDonald's franchisees as clients in a more competitive ecosystem will require proactive measures:

- **Loyalty Programs**: Taylor could introduce loyalty rewards for franchisees who continue using its services and equipment.

- **Training and Certification**: Offering comprehensive training programs for franchisee staff on machine operation and basic troubleshooting could reduce dependency on third-party services.

- **Technology Partnerships**: Collaborating with third-party innovators like Kytch to integrate advanced diagnostics and predictive maintenance features directly into Taylor machines may help mitigate the impact of independent repair solutions.

The DMCA exemption represents a significant disruption to Taylor's long-standing monopoly on McDonald's ice cream machine repairs. While it poses challenges in terms of revenue loss and increased competition, it also presents an opportunity for Taylor to rethink its business model, improve its product offerings, and strengthen relationships with franchisees. By adopting a strategic, customer-centric approach, Taylor can remain a key player in the commercial food equipment market.

4.5 Cultural and Regional Factors

Measuring performance across different cultural and regional environments presents significant challenges due to variations in local practices, consumer behaviors, and operational standards. These factors can directly impact machine performance, usage intensity, and service expectations, leading to disparities in performance metrics among distributors.

Key Regional Variations and Their Impacts

- **Variation in Product Serving Sizes**
  Serving sizes vary significantly across different regions, directly impacting machine strain and the frequency of breakdowns. For example, in Germany, ice cream cafes such as Café Brezels have adopted a pricing model that charges customers by the gram rather than by a fixed scoop size, allowing for more flexible serving portions and mitigating excessive machine use during peak hours (SCMP, 2024). In contrast, regions with larger standard serving sizes, such as the United States where scoops typically range from 3 to 5 ounces (KitchenJournal, 2024), place higher operational demands on machines, potentially accelerating wear and tear (Ashurst, 2024).
- **Cultural Preferences and Demand Peaks**
  Cultural factors influence portion sizes and consumption patterns. In Mediterranean countries like Italy, where smaller servings of artisanal gelato are common, machines may require less frequent maintenance due to reduced strain (Ashurst, 2024). Conversely, in locations where high-volume orders or multi-scoop servings are prevalent, such as urban U.S. areas during the summer season, the increased workload can lead to higher failure rates and more frequent service calls (IceCreamBoss, 2024).
- **Operational Practices and Staff Training Levels**
  The level of staff training and adherence to machine operation protocols can vary across regions. In some areas, insufficient training on proper machine handling may lead to improper usage, increasing breakdown frequency and repair needs. Conversely, regions with well-trained staff may exhibit better machine longevity and lower maintenance costs.

Disparities in Performance Metrics

Cultural and regional differences can lead to significant disparities in key performance indicators (KPIs), such as breakdown rate, response time, and service coverage. Without proper adjustments, distributors operating in more demanding environments may appear less efficient, resulting in unfair comparisons.

4.5.1 Proposed Methods for Adjusting Performance Evaluations

To ensure fairness and accuracy in assessments, distributors should adopt the following methods:

- **Weighted Metrics Adjustment**
  Introduce weighting factors for KPIs based on regional demand intensity, environmental conditions, and serving size norms. For instance, breakdown rates in high-demand regions can be normalized using a demand variability index derived from historical data.
- **Seasonal Performance Benchmarking**
  Compare distributors' performance during equivalent demand periods (e.g., summer months or local festival seasons) rather than using a uniform annual average. This approach ensures that performance is assessed relative to similar operational conditions.
- **Operational Context Normalization**
  Incorporate qualitative assessments of local operational practices, such as staff training programs and machine handling protocols, into the performance evaluation framework. Regions with less formalized training may require additional support rather than penalization for higher breakdown rates.

4.5.2 Strategies for Region-Specific Support and Training

To mitigate the impact of cultural factors on machine performance, Taylor Company can implement region-specific strategies:

- **Tailored Training Programs**
  Develop localized training modules that address specific operational challenges faced by staff in different regions. These modules should cover best practices for machine handling, common troubleshooting techniques, and preventive maintenance.

- **Proactive Maintenance Scheduling**
  In regions with high seasonal demand, Taylor should provide additional support by scheduling proactive maintenance checks before peak periods. This can reduce the likelihood of machine failures during critical times.

- **Localized Support Teams**
  Establish regional support teams equipped with knowledge of local cultural practices and operational environments. These teams can offer more responsive and context-aware service to franchisees.

- **Feedback-Driven Adaptation**
  Continuously gather feedback from distributors and franchisees on region-specific issues and incorporate it into training, support, and machine design improvements. This iterative approach ensures that Taylor remains responsive to evolving regional needs.

By acknowledging and addressing cultural and regional factors, Taylor Company can enhance the accuracy of performance evaluations, foster distributor satisfaction, and ensure more reliable service delivery across diverse markets.

5 Conclusion

This report has established a comprehensive Performance Evaluation Framework for Taylor Company, focusing on the assessment and enhancement of its regional distributors responsible for maintaining McDonald's ice cream machines. By defining and analyzing key performance indicators (KPIs) such as response time, breakdown rate, service coverage, and proximity, the study identified critical areas influencing distributor efficiency and reliability. The integration of external data sources, such as weather patterns provided valuable context, revealing how environmental factors impact machine performance and distributor operations.

The proposed bonus allocation framework offers a structured approach to incentivize distributors based on their performance metrics, ensuring fairness and motivating continuous improvement. Strategic recommendations, including operational enhancements like route optimization and workload rebalancing, alongside leveraging external data for demand forecasting, are poised to significantly enhance service reliability and operational efficiency. Additionally, adapting to recent legal developments by diversifying product offerings and strengthening customer support will help Taylor Company navigate the evolving competitive landscape effectively.

In summary, by adopting a data-driven and strategic approach, Taylor Company can optimize its distributor network, minimize machine downtime, and sustain high levels of customer satisfaction. Implementing the recommended strategies will not only address the current challenges but also position Taylor Company for long-term growth and resilience in a dynamic market environment.

References

1. Hauptman, M. (2024). More McFlurrys: US Copyright Office allows McDonald's to fix broken ice cream machines. *USA Today*. Retrieved from `https://www.usatoday.com`

2. Wade-Vicente, A. (2022). Taylor's C602 McDonald's ice cream machines and the right to repair. *Richmond Journal of Law and Technology*. Retrieved from `https://www.richmondjolt.com`

3. Shepherd, C. (2024). McDonald's ice cream machines get help from the feds. *Axios*. Retrieved from `https://www.axios.com`

4. Ashurst, R. (2024). *The scoop on ice cream: Uncovering the standard scoop size*. KitchenJournal. Retreived from `kitchenjournal.net`

5. Husfarm. (2024). *The sweet science of ice cream pricing in the dairy market*. Husfarm. Retreived from `husfarm.com`

6. IceCreamBoss. (2024). *Frozen treats, flexible prices: Strategies for ice cream inventory pricing*. Ice Cream Business Boss. Retreived from `icecreambusinessboss.com/`

7. SCMP. (2024). Ice cream scoops priced per gram? Cafe owner says it makes summer treat more affordable. *South China Morning Post*.

6 Appendix

6.1 Detailed Results

6.1.1 Table A: Performance Metrics per Distributor

matched_stockist_id	distributor_name	service_coverage	total_breakdowns	breakdown_rate	avg_distance_km	breakdown_frequency	workload_ratio	composite_efficiency_kpi
21084323	DSL Northwest, Inc.	143	53401	3674.380734	72.71836173	373.4335664	3.042553191	0.3451008661
21084324	Taylor Freezers of California	239	64570	4442.889908	70.59852425	270.167364	5.085106383	0.214518731
21084325	DSL Northwest, Inc.	194	80941	5569.334862	64.59281393	417.2216495	4.127659574	0.2601464471
21084326	Taylor Freezers of California	199	63466	4366.926606	87.13960281	318.9246231	4.234042553	0.2507922801
21084327	Taylor Freezers of California	573	205473	14138.05046	52.03774098	358.591623	12.19148936	0.1040299423
21084328	Taylor Freezers of California	131	49778	3425.091743	57.98173381	379.9847328	2.787234043	0.3786571216
21084329	Taylor Freezers of California (Nevada)	115	37431	2575.527523	38.7598532	325.4869565	2.446808511	0.4375678632
21084332	Taylor Freezer Sales of Arizona	245	121776	8379.082569	69.71924014	497.044898	5.212765957	0.2081918683
21084333	Taylor Mountain West	106	39465	2715.481651	53.83011884	372.3113208	2.255319149	0.4646591113
21084334	Taylor Company	140	32187	2214.701835	107.2928961	229.9071429	2.978723404	0.3493841494
21084335	Lane & McClain Distributors, Inc.	60	28335	1949.655963	60.78425338	472.25	1.276595745	0.8019024851
21084336	Rocky Mountains Distributing	184	45538	3133.348624	59.73196396	247.4891304	3.914893617	0.2762168193
21084340	Lane Equipment Company	124	45057	3100.252294	51.49464195	363.3629032	2.638297872	0.4012038238
21084341	Lane Equipment Company	84	34008	2340	104.049988	404.8571429	1.787234043	0.5716045817
21084342	Lane Equipment Company	107	45103	3103.417431	54.96939498	421.5233645	2.276595745	0.4598166254
21084343	Taylor of Oklahoma	202	91903	6323.600917	120.5375433	454.9653465	4.29787234	0.2431674073
21084345	Lane & McClain Distributors, Inc.	334	125168	8612.477064	56.81259392	374.754491	7.106382979	0.1609887076
21084346	McCormack Distributing Company	135	49629	3414.83945	168.5800642	367.6222222	2.872340426	0.3568002314
21084347	Lane Equipment Company	331	133448	9182.201835	50.62401576	403.1661631	7.042553191	0.164227795
21084348	TSBL Distributing	237	54886	3776.559633	86.09031138	231.5864979	5.042553191	0.2142459863
21084349	TFGroup LLC - Little Rock, AR	133	39338	2706.743119	140.7629782	295.7744361	2.829787234	0.3638685541
21084350	TFGroup LLC - Denham Springs, LA	96	33374	2296.376147	78.90881797	347.6458333	2.042553191	0.5051326789
21084352	Midwest Equipment Company	263	71497	4919.518349	97.27813607	271.851711	5.595744681	0.1926655026
21084353	TFGroup LLC	120	43072	2963.669725	96.01314691	358.9333333	2.553191489	0.4048679397
21084354	TFGroup LLC - Flowood, MS	64	16653	1145.848624	103.6060063	260.203125	1.361702128	0.7478701015
21084355	TFGroup LLC - Memphis, TN	125	49392	3398.53211	82.68277833	395.136	2.659574468	0.3906251912
21084358	Taylor Enterprises of Wisconsin, Inc.	204	39161	2694.56422	93.78333708	191.9656863	4.340425532	0.2462642962
21084359	ABS & Taylor Enterprises, Inc.	495	169427	11657.8211	58.66140793	342.2767677	10.53191489	0.1149180898
21084361	TFGroup LLC - Nashville, TN	165	54999	3784.334862	74.23683447	333.3272727	3.510638298	0.3013189413
21084362	Taylor Sales & Service Inc.	156	33641	2314.747706	100.4780821	215.6474359	3.319148936	0.3158716692
21084365	Taylor Freezer of Michigan, Inc.	151	60415	4156.995413	70.29330977	400.0993377	3.212765957	0.3279837624
21084366	Taylor Enterprises of Kentucky	220	79238	5452.155963	95.42822759	360.1727273	4.680851064	0.226891889
21084367	Taylor Freezer of Michigan, Inc.	50	8320	572.4770642	119.2514303	166.4	1.063829787	0.954395259
21084369	Taylor Freezer Sales Company of Georgia, Inc.	384	144429	9937.775229	97.24983607	376.1171875	8.170212766	0.1353373728
21084370	Institutional Commercial Equipment, Inc.	204	58509	4025.848624	125.6514722	286.8088235	4.340425532	0.2418373219
21084372	Taylor Freezer of Michigan, Inc.	382	134214	9234.908257	64.71673546	351.3455497	8.127659574	0.1413348038
21084374	Taylor Freezer Sales Company, Inc.	428	151082	10395.55046	108.4619292	352.9953271	9.106382979	0.1218658077
21084375	Kappus Company	518	169771	11681.49083	126.9006211	327.7432432	11.0212766	0.1016649417
21084377	Taylor Ultimate Services Company	285	102952	7083.853211	65.84779811	361.2350877	6.063829787	0.1828670977
21084384	Taylor Equipment Distributors, Inc.	655	219554	15106.92661	96.26331352	335.1969466	13.93617021	0.0851272194
21084385	Taylor Upstate	195	48742	3353.807339	121.3149682	249.9589744	4.14893617	0.2532693035
21084387	Taylor Products, Inc.	366	123964	8529.633028	73.28312729	338.6994536	7.787234043	0.1450134765
21084388	Taylor Products, Inc.	461	185845	12787.5	48.53703693	403.1344902	9.808510638	0.1250356628
21084391	Taylor Upstate	113	44227	3043.142202	91.61833858	391.3893805	2.404255319	0.4293990493
21084392	Taylor Freezer of Connecticut, Inc.	233	69184	4760.366972	57.57767296	296.9270386	4.957446809	0.2224524122
21084393	Taylor Freezer of New England, Inc.	300	108547	7468.830275	67.63789858	361.8233333	6.382978723	0.1742150571
39317961	Midwest Equipment Company (Indiana)	350	141028	9703.761468	107.911596	402.9371429	7.446808511	0.1460343356

Table 1: Performance metrics per distributor.

6.1.2 Table B: Breakdown Frequency by Weather Condition

Code	Weather Description	Report Count	Occurrences	Breakdown Frequency
0	Cloudy	690,893,440	1,897,514	364.10
1	Drizzle	35,913,040	98,086	366.14
2	Foggy	158,019,618	452,423	349.27
3	Freezing Drizzle	267,658	770	347.61
4	Freezing Rain	396,910	1,113	356.61
5	Heavy Drizzle	24,982,705	68,355	365.48
6	Heavy Rain	46,678,169	127,244	366.84
7	Heavy Snow	16,281,172	45,673	356.47
8	Light Drizzle	67,623,736	185,749	364.06
9	Light Freezing Drizzle	632,653	1,761	359.26
10	Light Freezing Rain	415,978	1,121	371.08
11	Light Rain	44,842,758	122,848	365.03
12	Light Showers	5,915,974	15,701	376.79
13	Light Snow	10,440,667	28,974	360.35
14	Light Snow Showers	10,193	33	308.88
15	Mainly Sunny	93,291,113	251,754	370.56
16	Partly Cloudy	88,778,292	240,010	369.89
17	Rain	70,425,074	191,881	367.02
18	Showers	21,939,336	59,495	368.76
19	Snow	37,909,283	107,628	352.23
20	Snow Showers	894	5	178.80
21	Sunny	181,102,892	473,568	382.42
22	Thunderstorm	83,599,568	224,343	372.64

Table 2: Breakdown Frequency by Weather Condition

6.2 Plots

6.2.1 Plot A: Breakdown Frequency Over Time

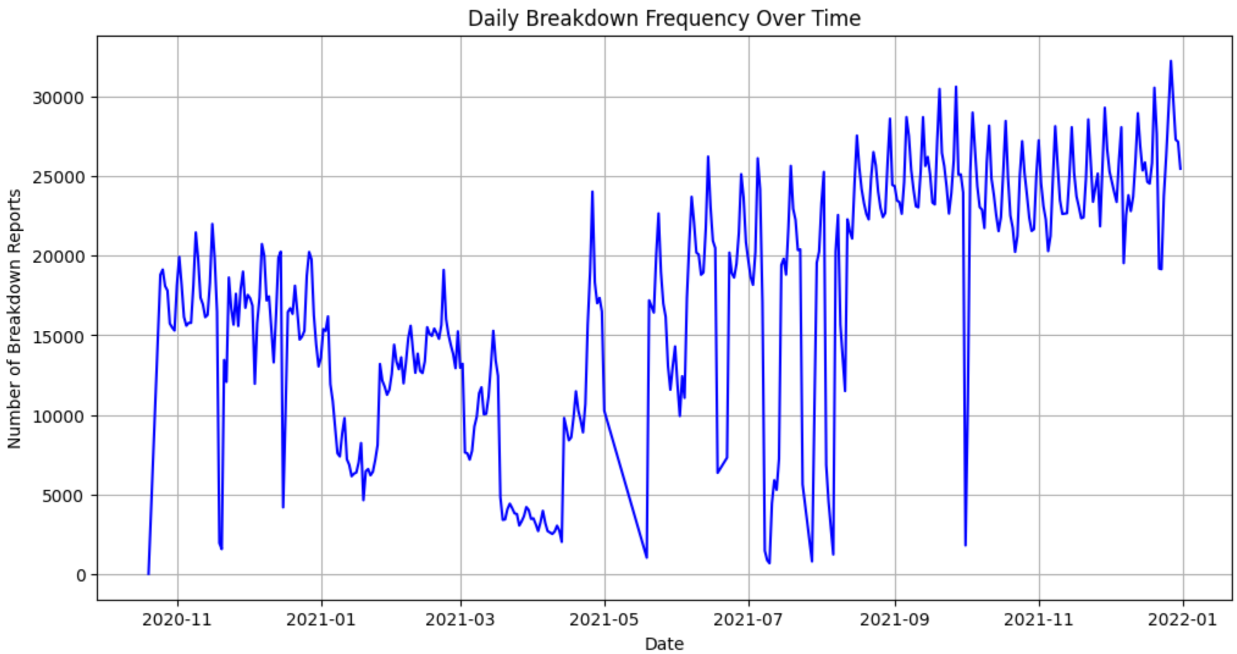

Figure 7: Breakdown Frequency Over Time

The high breakdown frequency observed after September can be attributed to a combination of seasonal and operational factors. As summer ends, many machines likely experience cumulative wear from peak usage during the warmer months, increasing the likelihood of failures. Additionally, seasonal transitions may involve fluctuating demand patterns, maintenance backlogs, or delays in routine servicing. External factors such as increased indoor activity and higher sales of warm desserts during cooler months could also contribute to the elevated machine workload, further driving up breakdown incidents. This pattern suggests a need for post-summer maintenance strategies to prevent performance drops.

6.2.2 Plot B: Breakdown Frequency by Weather Condition

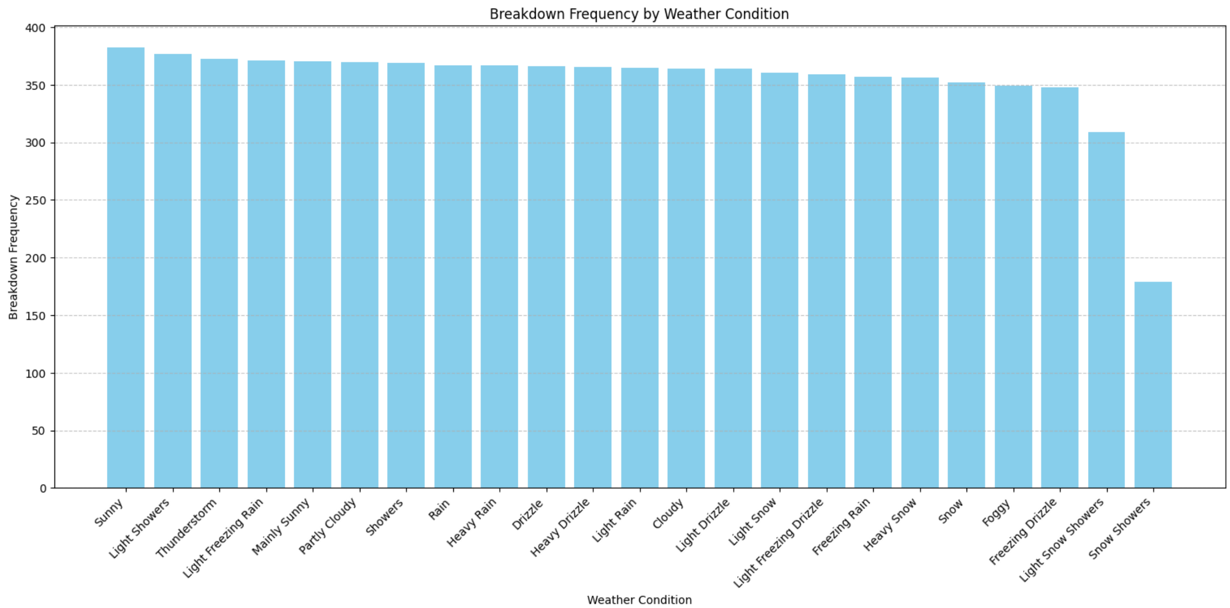

Figure 8: Breakdown Frequency by Weather Condition

6.2.3 Plot C: Correlation between Weather Condition and Breakdown Frequency

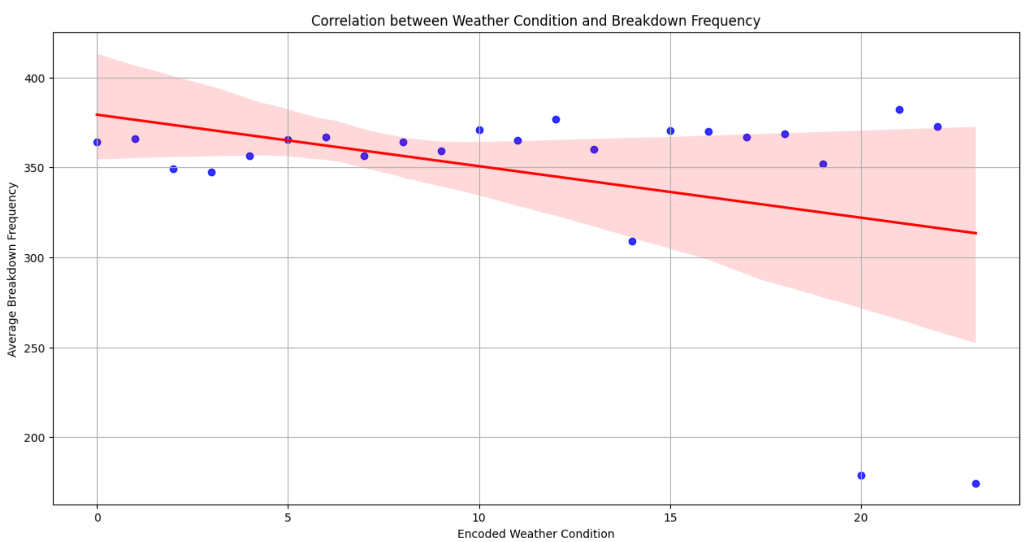

Figure 9: Correlation between Weather Condition and Breakdown Frequency

6.2.4 Plot D: Workload vs Efficiency

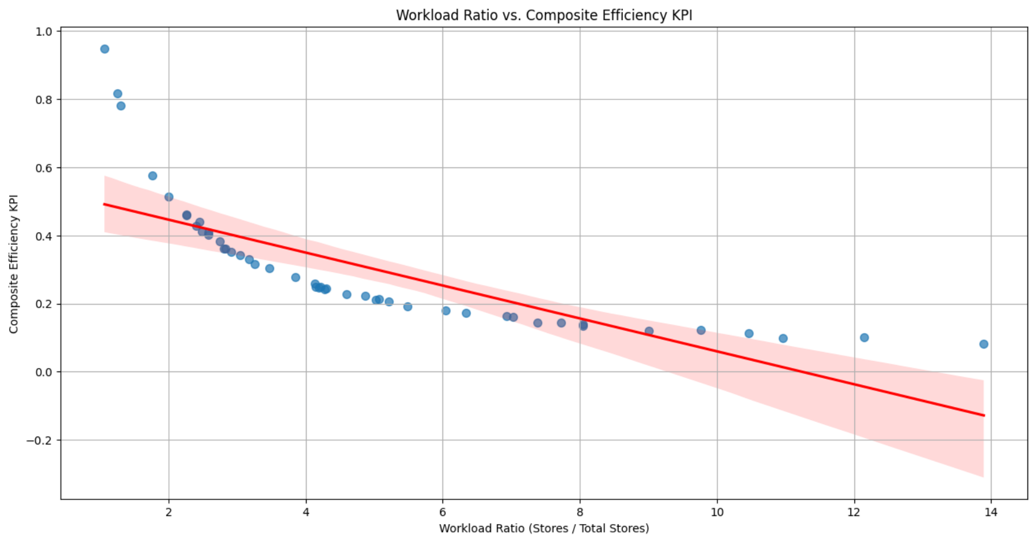

Figure 10: Workload vs Efficiency

6.2.5 Plot E: Efficiency vs Distance

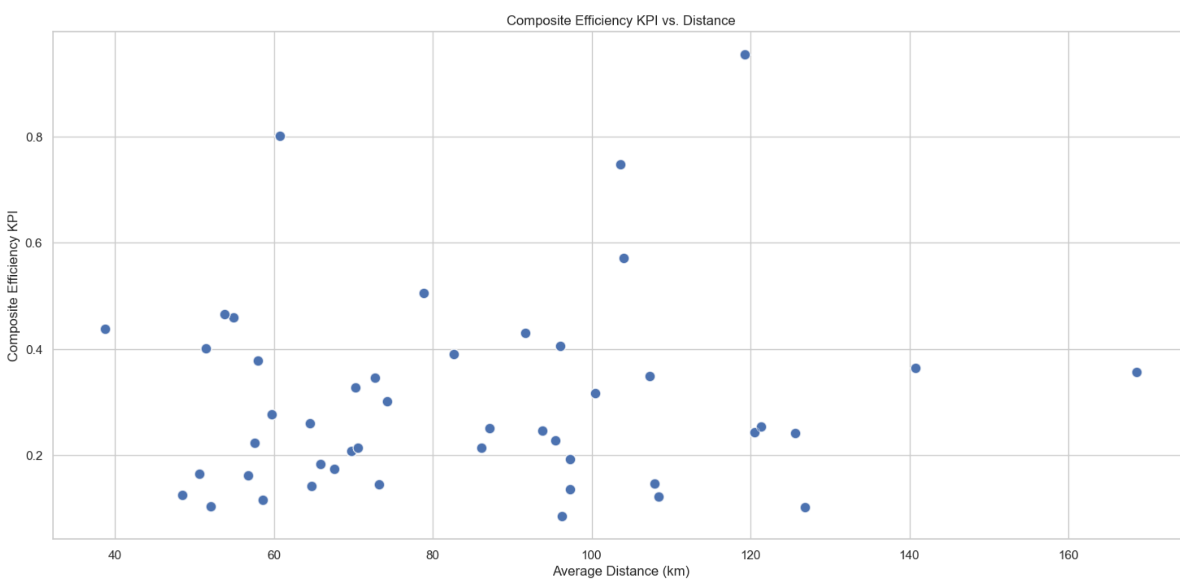

Figure 11: Efficiency vs Distance

6.2.6 Plot F: Breakdown Frequency vs Distance

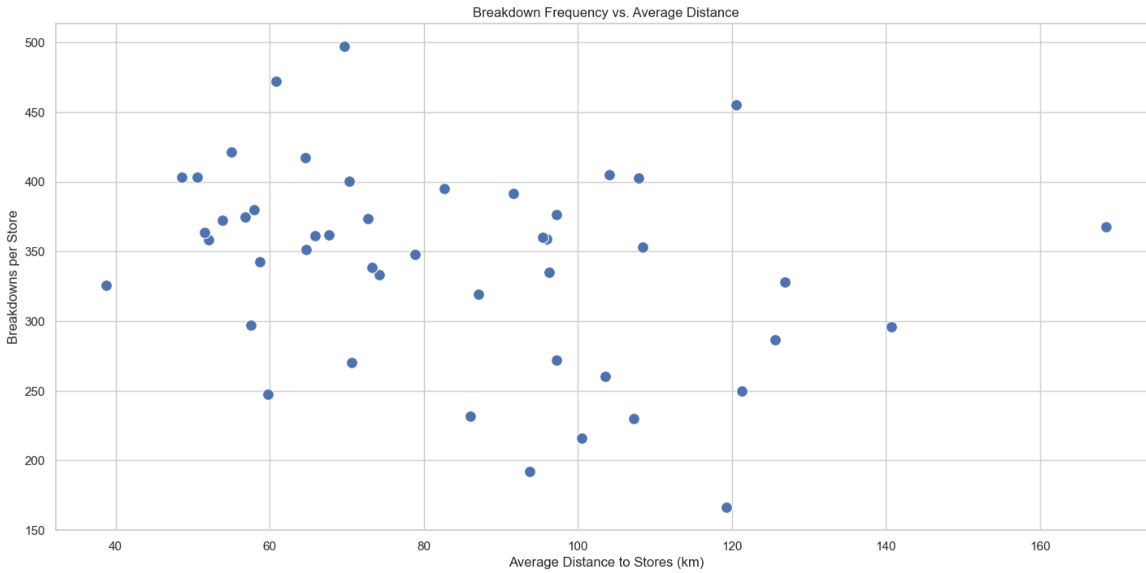

Figure 12: Breakdown Frequency vs Distance

6.2.7 Plot G: Breakdown Rate vs Distributor

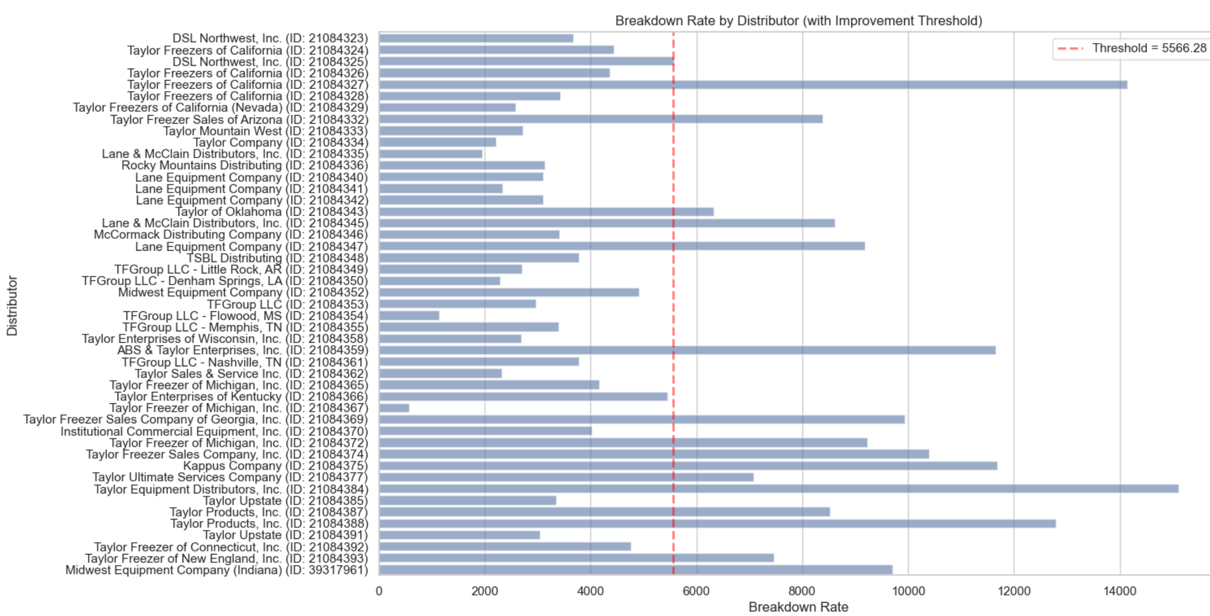

Figure 13: Breakdown Rate vs Distributor

For two of the plots, a threshold at the mean value was chosen for key performance indicators (KPIs) like the composite efficiency KPI and breakdown rate. The mean serves as a natural baseline for identifying underperforming distributors relative to the overall average, with scores below it flagged for improvement and those above it considered relatively efficient. This ensures an objective comparison and balances stringency with inclusiveness, preventing minor deviations from being overly penalized. Given the variability in distributor size, service coverage, and workload, the mean offers a fair, consistent benchmark. Alternative thresholds, such as one standard deviation below the mean, could be explored for more specific goals.

6.3 Code

6.3.1 Code A: Python Code for Weather Data Integration

Below is the Python code used to fetch and preprocess historical weather data from the Open Meteo API:

```python
import openmeteo_requests
import requests_cache
import pandas as pd
import time
from retry_requests import retry

# Setup the Open-Meteo API client with cache and retry on error
cache_session = requests_cache.CachedSession('.cache', expire_after=-1)
retry_session = retry(cache_session, retries=5, backoff_factor=0.2)
openmeteo = openmeteo_requests.Client(session=retry_session)

# Define the start date, end date, and other common parameters
start_date = "2020-10-01"
end_date = "2021-12-31"
daily_params = ["weather_code", "temperature_2m_max", "temperature_2m_min"]
timezone = "auto"

# Extract unique city locations
# Assuming 'city_locations' has columns: 'city', 'latitude', 'longitude'
latitude_list = unique_cities['latitude'].tolist()
longitude_list = unique_cities['longitude'].tolist()
city_list = unique_cities['city'].tolist()

# Load previously processed cities
try:
    processed_cities = pd.read_csv('final_weather_data_for_cities.csv')['city'].tolist()
    print(f"Resuming... {len(processed_cities)} cities already processed.")
except FileNotFoundError:
    processed_cities = []

# Prepare an empty list to store results
all_weather_data = []

url = "https://historical-forecast-api.open-meteo.com/v1/forecast"

# Loop through cities and fetch weather data for those not yet processed
for city, lat, lon in zip(city_list, latitude_list, longitude_list):
    if city in processed_cities:
        continue  # Skip cities already processed

    params = {
        "latitude": lat,
        "longitude": lon,
```

```python
        "start_date": start_date,
        "end_date": end_date,
        "daily": daily_params,
        "timezone": timezone
    }

    try:
        # Fetch the weather data for the current city
        responses = openmeteo.weather_api(url, params=params)
        response = responses[0]

        # Process daily data
        daily = response.Daily()
        daily_weather_code = daily.Variables(0).ValuesAsNumpy()
        daily_temperature_2m_max = daily.Variables(1).ValuesAsNumpy()
        daily_temperature_2m_min = daily.Variables(2).ValuesAsNumpy()

        # Create a DataFrame for the current city's weather data
        daily_data = {
            "date": pd.date_range(
                start=pd.to_datetime(daily.Time(), unit="s", utc=True),
                end=pd.to_datetime(daily.TimeEnd(), unit="s", utc=True),
                freq=pd.Timedelta(seconds=daily.Interval()),
                inclusive="left"
            ),
            "city": city,
            "weather_code": daily_weather_code,
            "temperature_2m_max": daily_temperature_2m_max,
            "temperature_2m_min": daily_temperature_2m_min
        }
        daily_dataframe = pd.DataFrame(data=daily_data)

        # Append the DataFrame to the results list
        all_weather_data.append(daily_dataframe)

        # Add city to processed list
        processed_cities.append(city)

        # Save the processed city list periodically
        pd.DataFrame({"city": processed_cities}).to_csv('processed_cities.csv', index=False)

        # Print progress
        print(f"Processed weather data for {city} ({lat}, {lon})")

    except Exception as e:
        # Log errors for the current city
        print(f"Failed to process weather data for {city} ({lat}, {lon}). Error: {e}")

    # Sleep to avoid exceeding API rate limits
    time.sleep(1)

# Read the existing data from the CSV
final_weather_data = pd.read_csv('final_weather_data_for_cities.csv')
```

```python
# Create a new DataFrame from the newly fetched data
new_weather_data = pd.concat(all_weather_data, ignore_index=True)

# Concatenate the existing data with the new data
final_weather_data = pd.concat([final_weather_data, new_weather_data], ignore_index=True)
print(final_weather_data.head())

# Save the final DataFrame to a CSV file
final_weather_data.to_csv('final_weather_data_for_cities.csv', index=False)
```

6.3.2 Code B: Python Code for Analysis and Visuals

Below is the Python code used to analyse the data and plot the graphs:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import folium
import numpy as np

# Load dataset
merged_data = pd.read_csv('merged_data.csv', low_memory=False)
pd.set_option('display.max_columns', None)

import pandas as pd

# merged_data = pd.read_csv('merged_data.csv')  # Load your actual dataset

# Define the weather code mapping
weather_code_mapping = {
    0: "Sunny",
    1: "Mainly Sunny",
    2: "Partly Cloudy",
    3: "Cloudy",
    45: "Foggy",
    48: "Rime Fog",
    51: "Light Drizzle",
    53: "Drizzle",
    55: "Heavy Drizzle",
    56: "Light Freezing Drizzle",
    57: "Freezing Drizzle",
    61: "Light Rain",
    63: "Rain",
    65: "Heavy Rain",
    66: "Light Freezing Rain",
    67: "Freezing Rain",
    71: "Light Snow",
    73: "Snow",
    75: "Heavy Snow",
    77: "Snow Grains",
    80: "Light Showers",
    81: "Showers",
    82: "Heavy Showers",
    85: "Light Snow Showers",
    86: "Snow Showers",
    95: "Thunderstorm",
```

```
    96: "Light Thunderstorms With Hail",
    99: "Thunderstorm With Hail"
}

# Map the weather codes to descriptions
merged_data['weather_description'] = merged_data['weather_code'].map(weather_code_mapping)

# Display the first few rows to verify
print(merged_data[['weather_code', 'weather_description']].head())


# Summary statistics
print(merged_data.shape)
print(merged_data.info())
print(merged_data.describe())

# Missing values
print('Missing values:', merged_data.isnull().sum())

# Convert 'mean_time_between_breakdowns' to timedelta and calculate total hours
merged_data['mean_time_between_breakdowns'] =
↪  pd.to_timedelta(merged_data['mean_time_between_breakdowns'], errors='coerce')

# Display summary statistics
print(merged_data.describe())

# Histograms for temperature
plt.figure(figsize=(10, 6))
merged_data['temperature_2m_max'].hist(bins=20, edgecolor='black')
plt.title('Histogram of Maximum Temperature')
plt.xlabel('Temperature (°C)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()

plt.figure(figsize=(10, 6))
merged_data['temperature_2m_min'].hist(bins=20, edgecolor='black')
plt.title('Histogram of Minimum Temperature')
plt.xlabel('Temperature (°C)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()

# Bar plot for weather_code
plt.figure(figsize=(10, 6))
merged_data['weather_code'].value_counts().plot(kind='bar', color='skyblue')
plt.title('Frequency of Weather Codes')
plt.xlabel('Weather Code')
plt.ylabel('Count')
plt.grid(True)
plt.show()

# Plot average daily max temperature over time
plt.figure(figsize=(12, 6))
merged_data.groupby('date')['temperature_2m_max'].mean().plot()
plt.title('Average Daily Maximum Temperature Over Time')
plt.xlabel('Date')
plt.ylabel('Temperature (°C)')
plt.grid(True)
```

```
plt.show()

import pandas as pd
import matplotlib.pyplot as plt

# Group by weather description and count the number of breakdowns
breakdown_by_weather =
↪   merged_data.groupby('weather_description')['report_count'].sum().reset_index()

# Calculate total occurrences of each weather description
weather_occurrences = merged_data['weather_description'].value_counts().reset_index()
weather_occurrences.columns = ['weather_description', 'occurrences']

# Merge the two datasets to get breakdowns and occurrences
weather_breakdown_analysis = pd.merge(breakdown_by_weather, weather_occurrences,
↪   on='weather_description')

# Calculate breakdown frequency (breakdowns per occurrence of weather condition)
weather_breakdown_analysis['breakdown_frequency'] = weather_breakdown_analysis['report_count'] /
↪   weather_breakdown_analysis['occurrences']

# Sort by breakdown frequency for better visualization
weather_breakdown_analysis = weather_breakdown_analysis.sort_values(by='breakdown_frequency',
↪   ascending=False)

# Plot the breakdown frequency per weather condition
plt.figure(figsize=(14, 8))
plt.bar(weather_breakdown_analysis['weather_description'],
↪   weather_breakdown_analysis['breakdown_frequency'], color='skyblue')
plt.xticks(rotation=45, ha='right')
plt.title('Breakdown Frequency by Weather Condition')
plt.xlabel('Weather Condition')
plt.ylabel('Breakdown Frequency')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

# Display the result
print(weather_breakdown_analysis)


import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder

# Encode weather descriptions to numeric values
label_encoder = LabelEncoder()
merged_data['weather_encoded'] = label_encoder.fit_transform(merged_data['weather_description'])

# Group by encoded weather and calculate average breakdown frequency
weather_breakdown_freq = (
    merged_data.groupby('weather_encoded'
    ['report_count'].mean().reset_index(name='average_breakdown_frequency')
)

# Calculate correlation between encoded weather and breakdown frequency
correlation = weather_breakdown_freq['weather_encoded'].
    corr(weather_breakdown_freq['average_breakdown_frequency'])
```

```python
print(f"Correlation between weather condition and breakdown frequency: {correlation:.2f}")

# Plot the correlation
plt.figure(figsize=(10, 6))
sns.regplot(
    x='weather_encoded',
    y='average_breakdown_frequency',
    data=weather_breakdown_freq,
    scatter_kws={'color': 'blue'},
    line_kws={'color': 'red'}
)
plt.title('Correlation between Weather Condition and Breakdown Frequency')
plt.xlabel('Encoded Weather Condition')
plt.ylabel('Average Breakdown Frequency')
plt.grid(True)
plt.show()
```

```python
# Box plots for temperature
plt.figure(figsize=(10, 6))
merged_data.boxplot(column=['temperature_2m_max', 'temperature_2m_min'], grid=False)
plt.title('Box Plot of Maximum and Minimum Temperatures')
plt.ylabel('Temperature (°C)')
plt.show()
```

```python
# Define a custom severity mapping based on weather codes
weather_severity_mapping = {
    0: 0,   # Sunny
    1: 0,   # Mainly Sunny
    2: 1,   # Partly Cloudy
    3: 1,   # Cloudy
    45: 2,  # Foggy
    48: 2,  # Rime Fog
    51: 2,  # Light Drizzle
    53: 3,  # Drizzle
    55: 3,  # Heavy Drizzle
    56: 4,  # Light Freezing Drizzle
    57: 4,  # Freezing Drizzle
    61: 3,  # Light Rain
    63: 4,  # Rain
    65: 5,  # Heavy Rain
    66: 5,  # Light Freezing Rain
    67: 6,  # Freezing Rain
    71: 3,  # Light Snow
    73: 4,  # Snow
    75: 5,  # Heavy Snow
    77: 3,  # Snow Grains
    80: 3,  # Light Showers
    81: 4,  # Showers
    82: 5,  # Heavy Showers
    85: 3,  # Light Snow Showers
    86: 4,  # Snow Showers
    95: 6,  # Thunderstorm
    96: 7,  # Light Thunderstorms With Hail
    99: 8   # Thunderstorm With Hail
```

```
}

# Apply the severity mapping to the dataset
merged_data['weather_severity'] = merged_data['weather_code'].map(weather_severity_mapping)

# Group by weather severity and calculate average breakdown frequency
severity_breakdown_freq = (
    merged_data.groupby('weather_severity')['report_count'].mean().
    reset_index(name='average_breakdown_frequency')
)

# Plot the relationship between weather severity and breakdown frequency
plt.figure(figsize=(10, 6))
plt.plot(severity_breakdown_freq['weather_severity'],
↪   severity_breakdown_freq['average_breakdown_frequency'], marker='o')
plt.title('Breakdown Frequency vs. Weather Severity')
plt.xlabel('Weather Severity (Encoded)')
plt.ylabel('Average Breakdown Frequency')
plt.grid(True)
plt.show()

import pandas as pd
import matplotlib.pyplot as plt

# Group by weather description and calculate the average breakdown frequency
weather_breakdown_freq = (
    merged_data.groupby('weather_description')['report_count'].mean().
    reset_index(name='average_breakdown_frequency')
)

# Sort the data by average breakdown frequency for better visualization
weather_breakdown_freq = weather_breakdown_freq.sort_values(by='average_breakdown_frequency',
↪   ascending=False)

# Plot the bar plot with descriptive weather names
plt.figure(figsize=(14, 8))
plt.bar(weather_breakdown_freq['weather_description'],
↪   weather_breakdown_freq['average_breakdown_frequency'], color='skyblue')
plt.xticks(rotation=45, ha='right')
plt.title('Average Breakdown Frequency by Weather Condition')
plt.xlabel('Weather Condition')
plt.ylabel('Average Breakdown Frequency')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# For nicer plotting style:
sns.set_theme(style="whitegrid")  # or "darkgrid", "ticks", etc.

final_distributor_df = pd.read_csv('distributor_kpis.csv')

final_distributor_df['distributor_label'] = (
    final_distributor_df['distributor_name']
    + " (ID: " + final_distributor_df['matched_stockist_id'].astype(str) + ")"
)
```

```python
# Sort distributors by composite_efficiency_kpi (descending) for easier viewing
df_sorted = final_distributor_df.sort_values(by='composite_efficiency_kpi', ascending=False)
'''
plt.figure(figsize=(12, 6))
sns.barplot(
    data=df_sorted,
    x='distributor_name',
    y='composite_efficiency_kpi',
    alpha=0.5
)

plt.xticks(rotation=90)  # Rotate distributor names if they overlap
plt.title('Composite Efficiency KPI by Distributor')
plt.xlabel('Distributor Name')
plt.ylabel('Composite Efficiency KPI')
plt.tight_layout()
plt.show()'''


plt.figure(figsize=(8, 6))
sns.scatterplot(
    data=final_distributor_df,
    x='avg_distance_km',
    y='breakdown_frequency',
    s=100,  # marker size
    alpha=0.6,
    legend=False  # remove the legend
)

plt.title('Breakdown Frequency vs. Average Distance')
plt.xlabel('Average Distance to Stores (km)')
plt.ylabel('Breakdowns per Store')
plt.tight_layout()
plt.show()



# Subset columns of interest
kpi_cols = [
    'service_coverage',
    'total_breakdowns',
    'breakdown_rate',
    'avg_distance_km',
    'breakdown_frequency',
    'workload_ratio',
    'composite_efficiency_kpi'
]

corr_matrix = final_distributor_df[kpi_cols].corr()

plt.figure(figsize=(8, 6))
sns.heatmap(
    corr_matrix,
    annot=True,
    cmap='Blues',
    alpha=0.6,
```

```python
    fmt='.2f',
    vmin=-1,
    vmax=1
)
plt.title('Correlation Heatmap of KPIs')
plt.show()



# read the McBroken Data (identified by store_id)
McDonalBrokensUS = pd.read_csv(
    'https://www.dropbox.com/scl/fi/3zf4vlp5j74ymjawe86tu/
    McDonadlBrokensUS.csv?rlkey=1gs1frdkbbiaiv4rws47u6mws&dl=1'
)
# read McDonald Branches Data and matched with a Taylor Retailer (identified by store_id and
↪  Stockist ID)
McDonalStoresUS = pd.read_csv(
    'https://www.dropbox.com/scl/fi/6ll8iu44hzlj7la5moaum/
    McDonalStoresUS.csv?rlkey=s7s6yavcwhr1bhlj0uo5kajy8&dl=1'
)
# read Taylor Retailers (identified by Stockist ID) information:
TaylorUS = pd.read_csv(
    'https://www.dropbox.com/scl/fi/ydzy3b8n1lnk84a9fxww3/
    TaylorUS.csv?rlkey=14l111k6hewtb9it75vesl4a3&dl=1'
)

# Convert timestamp to datetime format
McDonalBrokensUS['timestamp'] = pd.to_datetime(McDonalBrokensUS['timestamp'])

# Sort data by store_id and timestamp
McDonalBrokensUS = McDonalBrokensUS.sort_values(by=['store_id', 'timestamp'])

# Calculate time difference between consecutive breakdowns for each store
McDonalBrokensUS['time_diff'] = McDonalBrokensUS.groupby('store_id')['timestamp'].diff()

# Identify and remove duplicate breakdowns (time diff < 30 minutes)
McDonalBrokensUS = McDonalBrokensUS[
    (McDonalBrokensUS['time_diff'].isna()) | (McDonalBrokensUS['time_diff'] >=
    ↪  pd.Timedelta(minutes=30))
]

# Step 1: Merge store-level (McDonalStoresUS) with final_distributor_df on matched_stockist_id
store_level_kpi_df = pd.merge(
    McDonalStoresUS,            # each row = a store, columns include 'state',
    ↪  'matched_stockist_id', etc.
    final_distributor_df,       # aggregator, one row per matched_stockist_id
    on='matched_stockist_id',   # common key
    how='left'
)

kpi_by_state = (
    store_level_kpi_df
    .groupby('state', as_index=False)['composite_efficiency_kpi']
    .mean()   # or .median(), .sum(), etc., depending on your analysis
    .rename(columns={'composite_efficiency_kpi': 'avg_efficiency_kpi'})
)

# Sort descending for a nicer bar chart
kpi_by_state.sort_values('avg_efficiency_kpi', ascending=False, inplace=True)
```

```
plt.figure(figsize=(24, 6))
sns.barplot(
    data=kpi_by_state,
    x='state',
    y='avg_efficiency_kpi',
    alpha=0.5
)
plt.title('Average Composite Efficiency KPI by State')
plt.xlabel('State')
plt.ylabel('Avg Efficiency KPI')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()

'''kpi_by_state_distributor = (
    store_level_kpi_df
    .groupby(['state', 'distributor_name'], as_index=False)['composite_efficiency_kpi']
    .mean()
    .rename(columns={'composite_efficiency_kpi': 'avg_efficiency_kpi'})
)

plt.figure(figsize=(12, 6))
sns.barplot(
    data=kpi_by_state_distributor,
    x='state',
    y='avg_efficiency_kpi',
    hue='distributor_name',
    palette='tab10'
)
plt.title('Average Efficiency KPI by State & Distributor')
plt.xlabel('State')
plt.ylabel('Avg Efficiency KPI')
plt.xticks(rotation=45, ha='right')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()'''

plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=store_level_kpi_df,
    x='avg_distance_km',
    y='composite_efficiency_kpi',
    alpha=0.6,
    s=80,
    legend=False
)
plt.title('Composite Efficiency KPI vs. Distance')
plt.xlabel('Average Distance (km)')
plt.ylabel('Composite Efficiency KPI')
plt.tight_layout()
plt.show()


heatmap_df = (
    store_level_kpi_df
    .groupby(['state', 'distributor_name'], as_index=False)['composite_efficiency_kpi']
    .mean()
    .pivot(index='state', columns='distributor_name', values='composite_efficiency_kpi')
```

```
)

plt.figure(figsize=(12, 8))
sns.heatmap(
    heatmap_df,
    annot=True,     # show the numeric value
    cmap='coolwarm',
    fmt='.2f'
)
plt.title('Heatmap of Average Efficiency KPI by State and Distributor')
plt.xlabel('Distributor Name')
plt.ylabel('State')
plt.tight_layout()
plt.show()'''


plt.figure(figsize=(10, 6))
bubble_sizes = final_distributor_df['workload_ratio'] * 100  # Scale bubble sizes for better
↪  visibility
sns.scatterplot(
    data=final_distributor_df,
    x='service_coverage',
    y='breakdown_rate',
    size='workload_ratio',
    sizes=(50, 500),
    alpha=0.6
)
plt.title('Breakdown Rate vs. Service Coverage (Bubble Size = Workload Ratio)')
plt.xlabel('Service Coverage (Number of Stores)')
plt.ylabel('Breakdown Rate')
plt.grid(True)
plt.tight_layout()
plt.show()

# Sort data by breakdown rate in descending order
final_distributor_df_sorted = final_distributor_df.sort_values(by='breakdown_rate',
↪  ascending=False)

plt.figure(figsize=(12, 24))
sns.barplot(data=final_distributor_df, x='breakdown_rate', y='distributor_label', alpha=0.6)

# Set a threshold value
threshold_value = final_distributor_df['breakdown_rate'].mean()

# Add a vertical line at the threshold
plt.axvline(x=threshold_value, color='red',alpha=0.5, linestyle='--', linewidth=2,
↪  label=f'Threshold = {threshold_value:.2f}')

# Add labels, title, and legend
plt.title('Breakdown Rate by Distributor (with Improvement Threshold)')
plt.xlabel('Breakdown Rate')
plt.ylabel('Distributor')
plt.grid(True, axis='x')
plt.legend(loc='upper right')
plt.tight_layout()
plt.show()
```

```python
'''sns.pairplot(final_distributor_df, vars=['breakdown_frequency', 'service_coverage', 'avg_distance_km', 'workload_ratio'], diag_kind='kde', plot_kws={'alpha':0.7})
plt.suptitle('Pair Plot of Key Distributor Metrics', y=1.02)
plt.show()'''


# Sort data by breakdown rate in descending order
final_distributor_df_sorted = final_distributor_df.sort_values(by='composite_efficiency_kpi', ascending=True)

plt.figure(figsize=(24, 8))

# Plot the bar chart
sns.barplot(
    data=final_distributor_df_sorted,
    x='distributor_label',
    y='composite_efficiency_kpi',
    alpha=0.6
)

# Set the threshold value (mean of composite efficiency KPI)
threshold_value = final_distributor_df_sorted['composite_efficiency_kpi'].mean()

# Add a horizontal line at the threshold
plt.axhline(y=threshold_value, color='red', linestyle='--', linewidth=2, label=f'Threshold = {threshold_value:.2f}')

# Add labels, title, legend, and rotate x-ticks
plt.xticks(rotation=90)
plt.title('Composite Efficiency KPI by Distributor (with Threshold Line)')
plt.xlabel('Distributor Name')
plt.ylabel('Composite Efficiency KPI')
plt.legend(loc='upper right')
plt.tight_layout()
plt.show()


plt.figure(figsize=(12, 24))  # Adjust figure size (height is greater for horizontal plot)

# Plot the horizontal bar chart
sns.barplot(
    data=final_distributor_df_sorted,
    y='distributor_label',
    x='composite_efficiency_kpi',
    alpha=0.5
)

# Set the threshold value (mean of composite efficiency KPI)
threshold_value = final_distributor_df_sorted['composite_efficiency_kpi'].mean()

# Add a vertical line at the threshold
plt.axvline(x=threshold_value, color='red', linestyle='--', linewidth=2, label=f'Threshold = {threshold_value:.2f}')

# Add labels, title, legend
plt.title('Composite Efficiency KPI by Distributor (with Threshold Line)')
plt.xlabel('Composite Efficiency KPI')
plt.ylabel('Distributor Name')
plt.legend(loc='upper right')
```

```
plt.tight_layout()
plt.show()
```

The MTBB (Mean Time Between Breakdowns) was calculated by:

Computing the time difference (time_diff) between consecutive breakdown events for each store. Taking the mean of these time intervals for each store to get the average time between breakdowns. Storing this result as mean_time_between_breakdowns, representing how frequently breakdowns occur on average for each store.