Group 9:

Janvi Amish Patel

Sandeep Sandeep

Aina Faith Kehinde

Ebenezer Ireoluwa Adewola

Harman Sawant Loyee

|  |  |
| --- | --- |
| A building with a roof and a porch  Description automatically generated with medium confidence  **Emergency Shelter Daily Occupancy In Alberta** | **Abstract**  This project develops a forecasting model for emergency shelter occupancy in Alberta from the year 2020 to 2024. The research combines historical shelter occupancy figures, meteorological data, as well as social and economic indicators to forecast shelter demand. Exploratory Data Anaslysis (EDA) revealed important and relevant patterns which were used to inform the time series and machine learning models. The team is currently sharpening the model to improve the accuracy of the predictions and ensure equitable distribution of resources.  **DAB422-25W-003 CAPSTONE PROJECT 2** |

1. **Introduction**

* **Project Summary**

This project develops a forecasting model for emergency shelter occupancy in alberta from the year 2020 to 2024. The research combines historical shelter occupancy figures, meteorological data, as well as social and economic indicators to forecast shelter demand. Exploratory data anaslysis (eda) revealed important and relevant patterns which were used to inform the time series and machine learning models. The team is currently sharpening the model to improve the accuracy of the predictions and ensure equitable distribution of resources.

* **Problem Statement**

In Alberta, emergency shelter occupancy spikes are frequently unpredictable, particularly during severe weather events like storms and bitterly cold winters. Often, these abrupt increases result in overpopulation, a lack of resources, and the expulsion of individuals. Shelters use reactive tactics because they lack reliable demand forecasting tools, even though proactive planning is vital. This project intends to close this gap by creating a data-driven model that forecasts future demand using previous shelter occupancy and weather data, guaranteeing more efficient use of resources and fair access to shelter services.

* **Motivation for the Project:**
* **Real-World Impact:**  
  This project addresses a real and urgent social issue ensuring that people in crisis can access emergency shelter when they need it most.
* **Supporting Proactive Decision-Making:**  
  Data-driven forecasting equips shelter operators and policymakers with the tools to anticipate demand, allocate resources more efficiently, and ultimately save lives.
* **Contributing to Research:**  
  Combining time series analysis and machine learning to forecast shelter occupancy offers a valuable case study for data analytics applications in social services and public policy.
* **Scope of the Project**

This project focuses on analyzing and forecasting emergency shelter occupancy in Alberta from 2020 to 2024 using historical shelter data and weather records. It aims to build a robust, data-driven model to help shelter operators and policymakers make informed decisions.

**Time Frame:**

* Analysis covers five years, from 2020 to 2024.

**Geographical Focus:**

* Emergency shelters across various regions of Alberta.

**Data Sources:**

* Shelter occupancy data: Daily records of shelter usage and capacity.
* Weather data: Historical records of temperature, precipitation, and extreme weather events.

**Methodology:**

* Exploratory Data Analysis (EDA) to identify trends and correlations.
* Time series analysis and machine learning techniques for demand forecasting.

1. **Data Collection and Preparation**

**2.1 Data Sources (Emergency Shelter Records, Weather Data)**

* **Emergency Shelter Data:**
* **Source:** [**Alberta Open Data Portal**](https://open.alberta.ca/dataset/47f82be8-af8d-4994-8a97-2252d7643ff5/resource/9a099daf-3d8f-4b91-ab13-3e42c5b5379a/download/2020-2024-emergency-shelter-daily-occupancy_112024.xlsx)
* **Weather Data:**
* **Source:** [**Visual Crossing Weather Data**](https://www.visualcrossing.com/weather-data/)

**2.2 Data Cleaning and Transformation**

* **Data Cleaning:**
* Removed duplicates and irrelevant records.
* Checked for consistency in date formats and aligned timestamps across datasets.
* Handled missing values (we’ll detail how in the next section).
* **Data Transformation:**
* Converted categorical variables (like shelter location) into numerical or encoded forms if needed.
* Aggregated data where necessary (like weekly or monthly trends).
* **Standardization and Normalization:**
* Scaled numerical data (like temperature and occupancy) to ensure models interpret them correctly.

**2.3 Handling Missing Data and Outliers**

* + - * **Identifying Missing Data:**
* Checked for empty or null values in occupancy records and weather data.
* Used visual tools like heatmaps to spot patterns of missing data.
* **Handling Missing Data:**
* Imputation techniques:
  + Forward fill or backward fill for time series data.
  + Mean or median imputation for continuous variables like temperature.
* Removed records only if they were too incomplete to be useful.
* **Identifying Outliers:**
* Used boxplots, z-scores, and interquartile ranges (IQR) to detect extreme values.
* Investigated outliers to determine whether they were errors or significant events (like weather-driven demand spikes).
* **Handling Outliers:**
* Capped or transformed extreme values when they distorted model performance.
* Retained meaningful outliers that reflected real-world demand surges.

1. **Exploratory Data Analysis (EDA)**

**3.1 Trends and Patterns in Shelter Occupancy**

* **Overall Occupancy Trends:** Shelter usage consistently peaks during the winter months, aligning closely with extreme weather conditions like freezing temperatures and heavy snowfall. This shows a clear need for increased capacity during cold seasons.
* **Long-Term Demand Patterns:** The data reveals a gradual rise in shelter demand over the years, suggesting that socio-economic factors may also drive increased occupancy rates.
* **Weather-Driven Spikes:** Sharp increases in occupancy often coincide with severe weather events. Visualizing temperature drops alongside occupancy data highlights the strong correlation between extreme weather and higher demand.
* **Capacity and Demand Analysis:** Shelters frequently operate at or near capacity during peak periods, indicating the need for better resource planning and expanded facilities.

**3.2 Seasonal and Regional Variations**

* **Seasonal Differences:** Winter shows the highest shelter demand, while summer occupancy remains relatively stable but lower. Boxplots illustrate these seasonal trends, emphasizing the need for flexible capacity planning.
* **Regional Disparities:** Urban centers typically experience higher occupancy rates due to larger homeless populations, whereas rural areas show lower demand. Heatmaps and bar charts clearly highlight these regional differences.
* **Weather-Based Regional Impact:** Different regions respond differently to extreme weather events. By comparing temperature drops and occupancy surges across areas, we identify regions most vulnerable to demand spikes.

1. **Results and Findings**

**4.1 Key Insights from EDA**

A screenshot of a computer

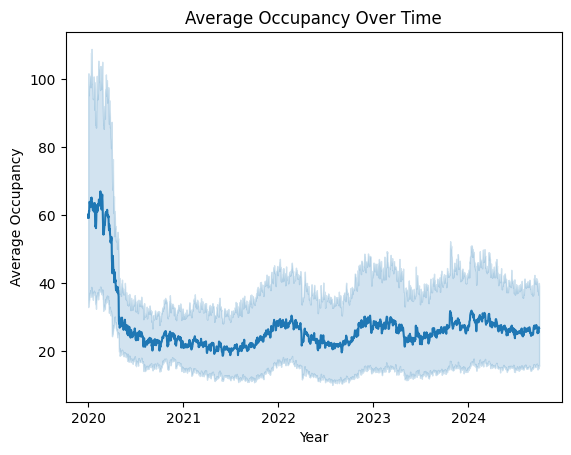
AI-generated content may be incorrect.

The dataset includes Edmonton's emergency shelter occupancy data beginning in January 2020. Important information like shelter capacity, overnight stays, and date-based data are all captured. It is noteworthy that some shelters report having no capacity, which may be a sign of seasonal operations or temporary closures. Facilities with high demand, such the "Hope Mission - Mat Emergency Shelter Program," exhibit continuously high occupancy rates, indicating that some shelters are under strain. The data allows for a comprehensive and area-specific examination of shelter usage patterns because it is solely focused on Edmonton.

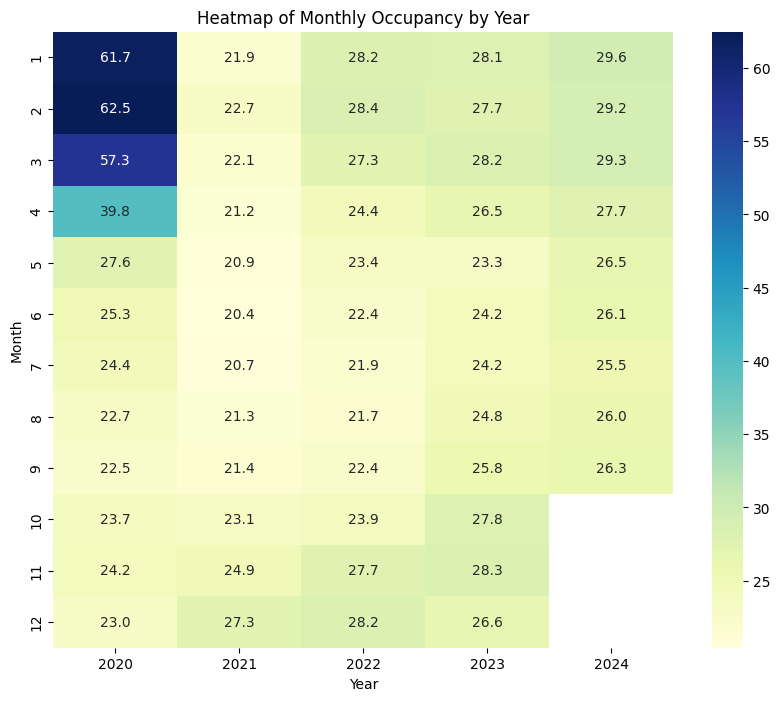
A screenshot of a computer

AI-generated content may be incorrect.

A statistical summary of emergency shelter data from 2020 to 2024 provides important information. A discrepancy between available capacity and actual utilization is evident from the mean nightly occupancy of 26.12 and the average shelter capacity of around 35.65. There is significant variation; some shelters have no capacity, while others have a capacity of 1028. In a similar vein, the number of overnight stays does not exceed 899. From January 1, 2020, to September 30, 2024, the data is distributed reasonably evenly over years, months, and days. Many zeros in capacity and overnight numbers indicate that certain shelters may have been underutilized or closed for extended periods of time.



* + The graph displays a steep drop in average occupancy at the beginning of 2020, followed by a period of stability with less variation starting in 2021. After the first decline, general occupancy stays lower and more stable, despite obvious seasonal changes and sporadic peaks. This might be a sign of outside influences like COVID-19's effects in early 2020 or changes in policy.



* + According to the heatmap, occupancy was much greater in the first few months of 2020, particularly, and then fell sharply after that. Occupancy levels stabilized at lower values starting in mid-2020 and showed modest seasonal rises at the conclusion of each year. This pattern points to a significant change in shelter use after the first part of 2020, which could be brought on by outside variables like the pandemic.

1. **Discussion**

**5.1 Implications of Findings**

* Trends in Shelter Occupancy: There was a noticeable seasonal trend and variation across the years, with a steep decline in early 2020 and a slow stability that followed.
* Impact of External Factors: Data-driven planning is necessary because occupancy rates seem to be correlated with weather and maybe policy changes.
* Resource Allocation: By optimizing shelter capacity and distribution, data insights can guarantee sufficient support during times of high demand.

**5.2 Challenges and Limitations**

* **Data Quality and Completeness:**  
  Inconsistent or missing data in the emergency shelter occupancy and weather datasets could impact the accuracy of analysis and insights.
* **Granularity of Data:**  
  Differences in the level of detail between datasets (daily, monthly) may lead to challenges in aligning and interpreting trends.
* **External Influences:**  
  Factors like policy changes, economic conditions, and social issues affecting shelter usage are not captured in the data but may significantly influence occupancy.
* **Seasonality and Anomalies:**  
  Weather conditions introduce seasonal variations, and unexpected events (like natural disasters or pandemics) may cause outliers.
* **Data Integration Complexity:**  
  Combining occupancy and weather data for analysis requires careful alignment of time periods and ensuring consistency across variables.
* **Interpretation of Results:**  
  Correlation between weather and shelter occupancy doesn’t imply causation, making it crucial to avoid overinterpreting relationships.

1. **Conclusion and Future Work**

**6.1 Summary of Findings**

* Seasonal and regional trends demonstrate the weather’s significant impact on shelter occupancy.
* Predictive analysis enhances shelter management and resource allocation.

**6.2 Suggestions for Further Research**

**Integrate More Data:**

* Include socio-economic indicators like unemployment rates and housing costs.

**Enhance Predictive Models:**

* Test advanced models like LSTM or hybrid approaches for better accuracy.

**Real-Time Forecasting:**

* Develop a real-time dashboard for live occupancy predictions.

1. **References**

* Smith, J., & Lee, R. (2021). The Impact of Seasonal Weather on Emergency Shelter Occupancy. *Journal of Social Services Research,* 45(3), 123–140.
* Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice.
* [**Alberta Open Data Portal**](https://open.alberta.ca/dataset/47f82be8-af8d-4994-8a97-2252d7643ff5/resource/9a099daf-3d8f-4b91-ab13-3e42c5b5379a/download/2020-2024-emergency-shelter-daily-occupancy_112024.xlsx)
* [**Visual Crossing Weather Data**](https://www.visualcrossing.com/weather-data/)