**Part 1: Temporal Analysis**

The dataset contains information on car crashes in Monroe County from 2003 to 2015. It includes variables that are useful for temporal analysis, such as:

* **Year, Month, Day, Hour**: Provides time-based granularity.
* **Weekend?**: Indicates whether the crash occurred on a weekend or weekday.
* **Collision Type**: Identifies the type of crash.
* **Injury Type**: Classifies injuries resulting from the crash.
* **Primary Factor**: Lists the main cause of the crash.
* **Reported Location, Latitude, Longitude**: Provides spatial information.

**Temporal Analysis Approach**

**Stakeholder Temporal Needs**

1. **Law Enforcement & Traffic Authorities**
   * Need to identify peak crash times (hourly/daily/seasonal trends).
   * Understand high-risk periods for different collision types.
   * Detect anomalies, such as unusual spikes in crashes.
   * Use findings for strategic deployment of officers and preventive measures.
2. **City Planners & Infrastructure Engineers**
   * Examine long-term crash trends to improve road design.
   * Compare seasonal patterns to adjust road maintenance plans.
   * Identify hazardous intersections for safety interventions.
3. **Public Safety & Emergency Services**
   * Need time-based insights to allocate emergency response resources efficiently.
   * Forecast high-crash periods to optimize ambulance and first responder readiness.
4. **Insurance Companies & Policy Makers**
   * Need yearly/monthly patterns to adjust insurance risk assessments.
   * Study the impact of new traffic laws on crash frequency.

**Data Assessment**

* **Time Representation**: Time series measurements (year, month, day, hour).
* **Temporal Granularity**: Hourly, daily, monthly, yearly.
* **Completeness**: Some missing values in the "Hour" and "Weekend?" columns.
* **Additional Context Needed**: Weather conditions, traffic volume, or road conditions could improve insights.

**Initial Visualization Exploration**

1. **Line Chart for Crash Trends Over Time**
   * X-axis: Time (Year/Month)
   * Y-axis: Number of crashes
   * Helps observe overall trends and seasonal variations.
2. **Heatmap of Hourly vs. Daily Crashes**
   * X-axis: Hour of the day
   * Y-axis: Day of the week
   * Color intensity: Number of crashes
   * Helps detect peak crash times (rush hours, late-night incidents).

A graph with blue lines

Description automatically generated

**Yearly Trend of Car Crashes (Line Chart)**

* + Shows fluctuations in crash occurrences over the years.
  + Identifies years with significant increases or decreases, which could be linked to policy changes, infrastructure improvements, or external factors.

A graph of a heatmap

Description automatically generated

**Heatmap of Crashes by Hour and Day of the Week**

* + Highlights peak crash times, such as rush hours (morning and evening).
  + Shows differences in crash distribution on weekdays vs. weekends.
  + Provides actionable insights for traffic authorities and emergency responders to allocate resources effectively.

**Part 2: AI-Assisted Design Process**

**Documenting AI Interactions**

**AI-Assisted Design Process**

**1. Documenting AI Interactions**

* **AI Model Used**: ChatGPT (GPT-4-turbo, January 2025 version)
* **Prompts Used**:
  1. *"How can I perform temporal analysis for a dataset containing car crash records with timestamps?"*
  2. *"What are some common temporal visualizations for accident data over time?"*
  3. *"Generate Python code for analyzing hourly and daily patterns in a dataset containing crash records with columns: Year, Month, Day, Hour, Collision Type, Injury Type, and Location."*
  4. *"How do I detect and visualize peak accident times in a dataset with timestamps?"*
  5. *"How can I handle missing time data in an accident dataset for effective visualization?"*
* **Why These Prompts?**
  1. Focused on **temporal analysis** to extract meaningful patterns.
  2. Emphasized **data preparation** and **visualization best practices**.
  3. Ensured solutions for **handling missing or incomplete data**.

**2. Implementation Plan**

**Data Preparation Steps**

* **Handling Missing Values**:
  + Fill missing **Hour** values with "Unknown" or use statistical imputation.
  + Drop incomplete rows only when necessary.
* **Convert Time Columns**:
  + Merge Year, Month, Day, Hour into a **datetime** object for analysis.
* **Categorize Data**:
  + Create a Day\_of\_Week column from the date.
  + Convert Hour into **time bins** (e.g., Morning, Afternoon, Evening, Night).
* **Data Aggregation**:
  + Summarize data by **hour, day, and season** to detect trends.

**Data Analysis and Visualization Tools**

* **Libraries**:
  + pandas for data processing
  + matplotlib for line plots and bar charts
  + seaborn for heatmaps and advanced visualizations
  + plotly (optional) for interactive charts
* **Why These Tools?**
  + **pandas**: Efficient time-series analysis.
  + **matplotlib/seaborn**: Well-suited for static visualizations.
  + **plotly**: Enables interactive exploration.

**Interactive Features**

* **Dropdown filters**: Allow stakeholders to filter by year, month, or crash type.
* **Hover tooltips**: Show crash details when hovering over data points.
* **Time slider**: Lets users explore crash trends dynamically.

**Handling Data Quality Issues**

* **Time gaps**: Fill in missing time periods with 0 crashes for consistency.
* **Anomalies**: Identify and remove outliers that skew trends.
* **Standardization**: Ensure time values are formatted uniformly.

**3. Evaluating AI Suggestions**

**Helpful AI Suggestions**

* **Recommended heatmaps** for visualizing hourly and daily crash trends.
* **Suggested aggregating data** by different time scales (daily, monthly, yearly).
* **Proposed data imputation techniques** to handle missing time values.

**Limitations Encountered**

* AI **did not account for seasonality** initially, so I had to manually include seasonal breakdowns.
* AI-generated visualizations **lacked annotations and clarity**, requiring improvements.
* **No interactive elements** were suggested, so I added plotly for user-friendly visuals.

**Modifications & Enhancements**

* Adjusted **color schemes** to enhance data visibility.
* Added **annotations** for peak accident times.
* Introduced **interactive components** for deeper analysis.

**Best Practices AI Missed**

* AI **suggested complex models too early**, but simpler **exploratory data analysis (EDA)** was more effective.
* AI didn’t **prioritize stakeholder needs**, so I ensured that visualizations directly answered practical questions.
* AI **missed data validation steps**, so I implemented checks to remove duplicate records and standardize time formats.

**Final Takeaways**

* AI **accelerated the ideation process**, but human refinement was necessary.
* **Combining AI-generated code with domain knowledge** led to better insights.
* **Best results came from iteratively modifying AI’s suggestions** rather than blindly following them.