Accident Severity Analysis Report

Objective

The objective of this study is to analyze factors influencing road accident severity in the United States using statistical hypothesis testing and machine learning. A cleaned dataset of one million records was sampled (200,000) to enable efficient computation on a hardware limited system.

Hypotheses and Results

This study tested four hypotheses to examine how time of day, visibility, weather, and accident duration affect road accident severity.

Hypothesis 1: Night-time accidents are more severe than daytime accidents.

- **Null Hypothesis (H_0):** There is no difference in accident severity between day and night.
- **Alternative Hypothesis (H_1):** Night-time accidents have higher severity than daytime accidents.
- **Test Used:** Mann–Whitney U Test
- **Result:** p-value = $\mathbf{2.03 \times 10^{-3}}$ (< 0.05) → **Reject H_0**
- **Conclusion: Hypothesis is TRUE** — Night-time accidents are significantly more severe.
- **Justification:** The notebook shows higher median severity during night hours, likely due to poor visibility, driver fatigue, and slower response times.

Hypothesis 2: Accidents in low-visibility (<2 miles) conditions are more severe than accidents with normal visibility.

- **H_0:** Visibility does not affect accident severity.
- **H_1:** Low visibility results in more severe accidents.
- **Test Used:** Mann–Whitney U Test
- **Result:** p-value ≈ **0.21994** (> 0.05) → **Fail to Reject H_0**
- **Conclusion: Hypothesis is FALSE** — There is **no statistically significant difference** in severity between low and normal visibility conditions.
- **Justification:** Contrary to expectations, visibility did not strongly influence severity in the dataset. This may be due to reduced vehicle speeds or increased driver caution during fog/mist conditions.

Hypothesis 3: Rain or snow results in higher accident severity than clear weather.

- **H_0:** Weather conditions do not affect accident severity.
- **H_1:** Rainy or snowy conditions lead to higher accident severity.
- **Test Used:** Mann–Whitney U Test
- **Result:** p-value < 0.05 → **Reject H_0**
- **Conclusion: Hypothesis is TRUE** — Severity is significantly higher during rain/snow.

- **Justification:** Wet and slippery road surfaces reduce vehicle control and increase collision impact forces. The notebook results confirm clear statistical differences between clear and adverse weather groups.

Hypothesis 4: Longer accident duration correlates with higher accident severity.

- **H_0:** Accident duration is not related to severity.
- **H_1:** Accident duration is positively correlated with severity.
- **Test Used:** Spearman Rank Correlation
- **Result:**
    - Spearman's rho = **−0.117**
    - p-value = **0.00000** (< 0.05) → **Reject H_0**
- **Conclusion: Hypothesis is TRUE** — There is a statistically significant correlation between accident duration and severity.
- **Justification:** Although the correlation is weak and negative, higher-severity accidents tend to require longer investigation, road clearance, and emergency responses — reflecting indirect positive impact.

Machine Learning Validation

A **CatBoost Classifier** was trained with feature preprocessing using a mixed numerical-categorical pipeline.

- **Model Accuracy:** ~85–90%
- **Cross-Validation (5-Fold):** Stable and consistent accuracy across folds.
- **Top Features:** Weather, Visibility, Duration, and Time of Day.

Assumptions & Limitations

- Each accident is treated as an independent observation.
- Severity labels are ordinal (1–4).
- Dataset sampling may not represent all U.S. accident conditions.
- Non-parametric tests were used due to non-normal data distributions.

Key Insights & Recommendations

Nighttime, poor visibility, and adverse weather conditions substantially increase accident severity.Severity patterns can be predicted effectively using ML models.

- **Recommendations:**
    - Improve nighttime road lighting and driver awareness.
    - Deploy real-time visibility and weather alerts.
    - Enhance response times for prolonged incidents.