1]Hasan Mhowwala [1]San Jose State University

# Discovering Patterns in Car Rental Data: An Empirical Analysis

[

September 30, 2023

**Abstract**

This research paper delves into the patterns hidden within a car rental dataset using the Knowledge Discovery in Databases (KDD) methodology. It illustrates the significance of data pre-processing, transformation, and mining techniques, highlighting how data-driven insights can be instrumental in predicting certain outcomes, such as car damages.

## 1 Introduction

The digitization of car rental processes has led to the accumulation of vast datasets, offering a myriad of opportunities for extracting actionable insights. This study focuses on a dataset comprising attributes such as pickup and dropoff locations, car type, payment methods, and reported damages upon car returns.

## 2 Methodology

The research employed the KDD methodology, comprising the following stages:

1. Data Understanding & Pre-processing
2. Data Cleaning
3. Data Transformation & Feature Engineering
4. Data Mining & Model Evaluation
5. Interpretation & Recommendations

## 3 Data Understanding & Pre-processing

The dataset consisted of 650 entries spanning across 10 attributes. Data types encompassing numerical, categorical, and date-time attributes. Upon visual exploration, distributions of attributes like popular pickup and dropoff locations, car types, and payment methods were discerned.

# 4 Data Cleaning

Despite the comprehensive nature of the dataset, data quality assurance remains pivotal. Our findings are as follows:

- Missing Values: The dataset demonstrated an absence of missing values.
- Outliers: Applying the Interquartile Range (IQR) method, anomalies were detected primarily within the 'Total_Distance' and 'Total_Amount' columns.

# 5 Data Transformation & Feature Engineering

To make the dataset conducive to Machine Learning models, numerical attributes were standardized, categorical variables underwent one-hot encoding, and new attributes were introduced. Visual examinations showcased discernible rental patterns.

# 6 Data Mining & Model Evaluation

The study revolved around predicting the propensity of damages being reported upon car return. Various classifiers were tested. Notably, the Decision Tree classifier was the most effective.

# 7 Interpretation & Recommendations

Key insights derived from the model include rentals of longer durations or extensive distances showcasing a heightened risk of damages and the influence of car type and payment methods on damage reports.

# 8 Conclusion

The study underscores the implications of data analytics in sectors such as car rentals. The patterns unveiled can significantly aid in proactive decision-making processes.

# 9 Acknowledgments

The researcher extends gratitude to all contributors and datasets providers.