



# Benford's Law Analysis of Crude Oil Dataset

Team Members:

1. Angelo Nelson – Data Preparation & Core Logic
2. Deepesh Dey – Data Analysis & Interpretation
3. Rohit Nair P – Research & Reporting
4. Sanket Jha – Presentation & Communication

# Introduction to Benford's Law

## Definition

Frequency distribution of first digits in data

$$P(d) = \log_{10}(1 + 1/d)$$

Applications

Fraud detection

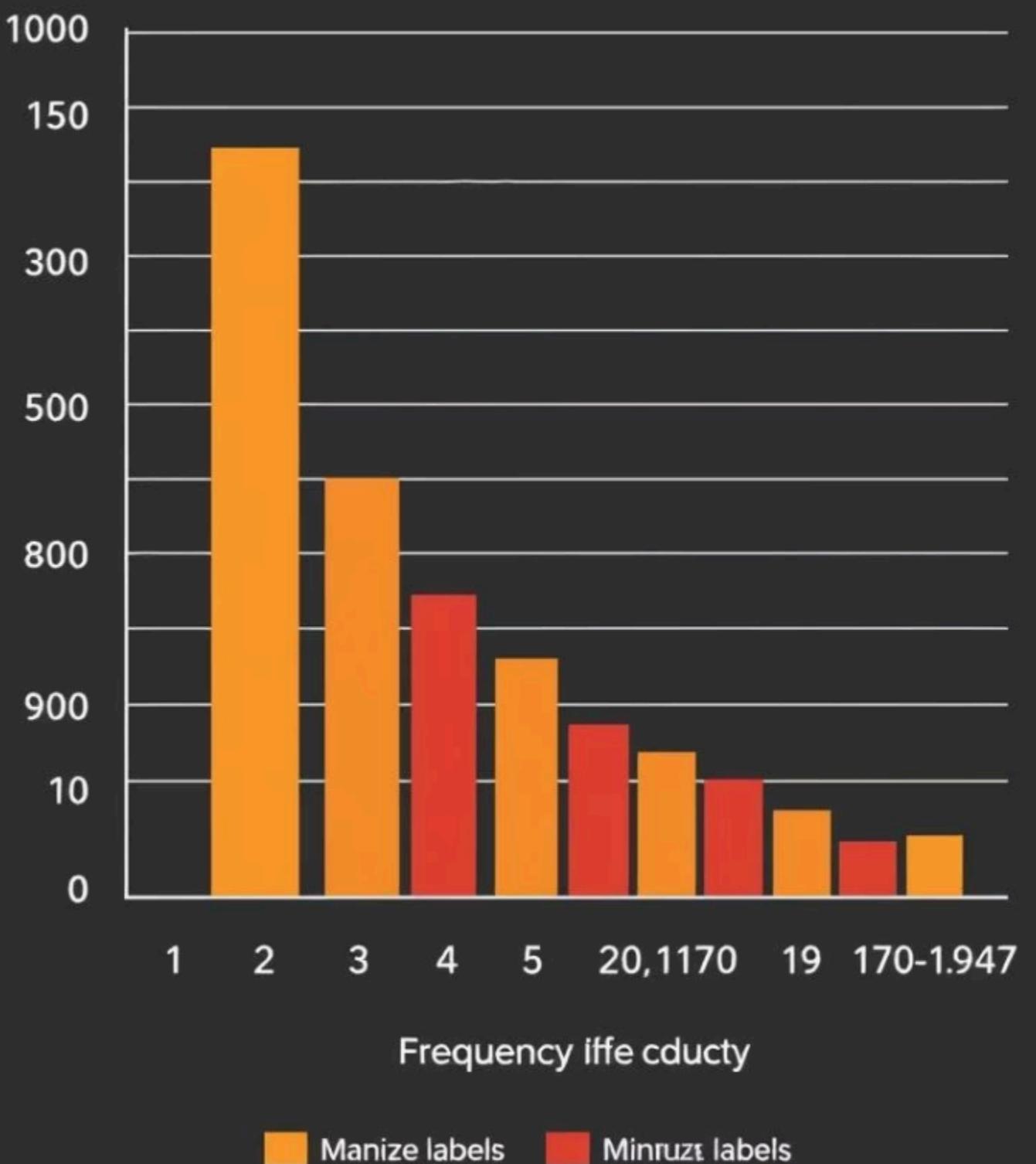
Auditing

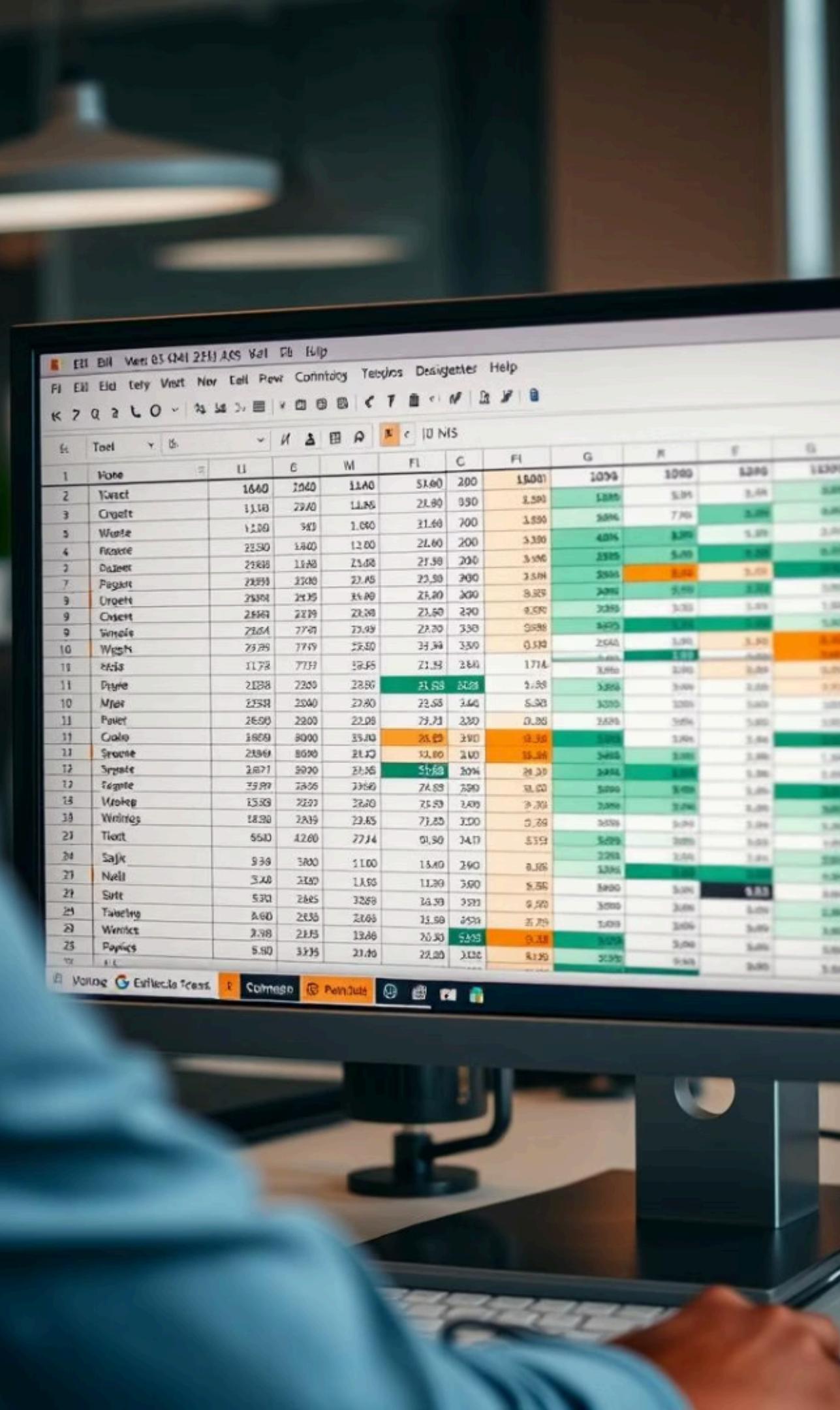
Anomaly checking

Digit Occurrence

Lower digits appear more frequently

# Benford Law





# Dataset Overview

## Source & Scope

Crude oil data covering multiple variables and time periods.

## Columns Analyzed

Value in Rupees, Value in Dollars, Quantity in Metric Tonnes.

# Data Cleaning and Preparation

- 1 Remove Duplicates
- 2 Handle Missing Values
- 3 Convert Data Types

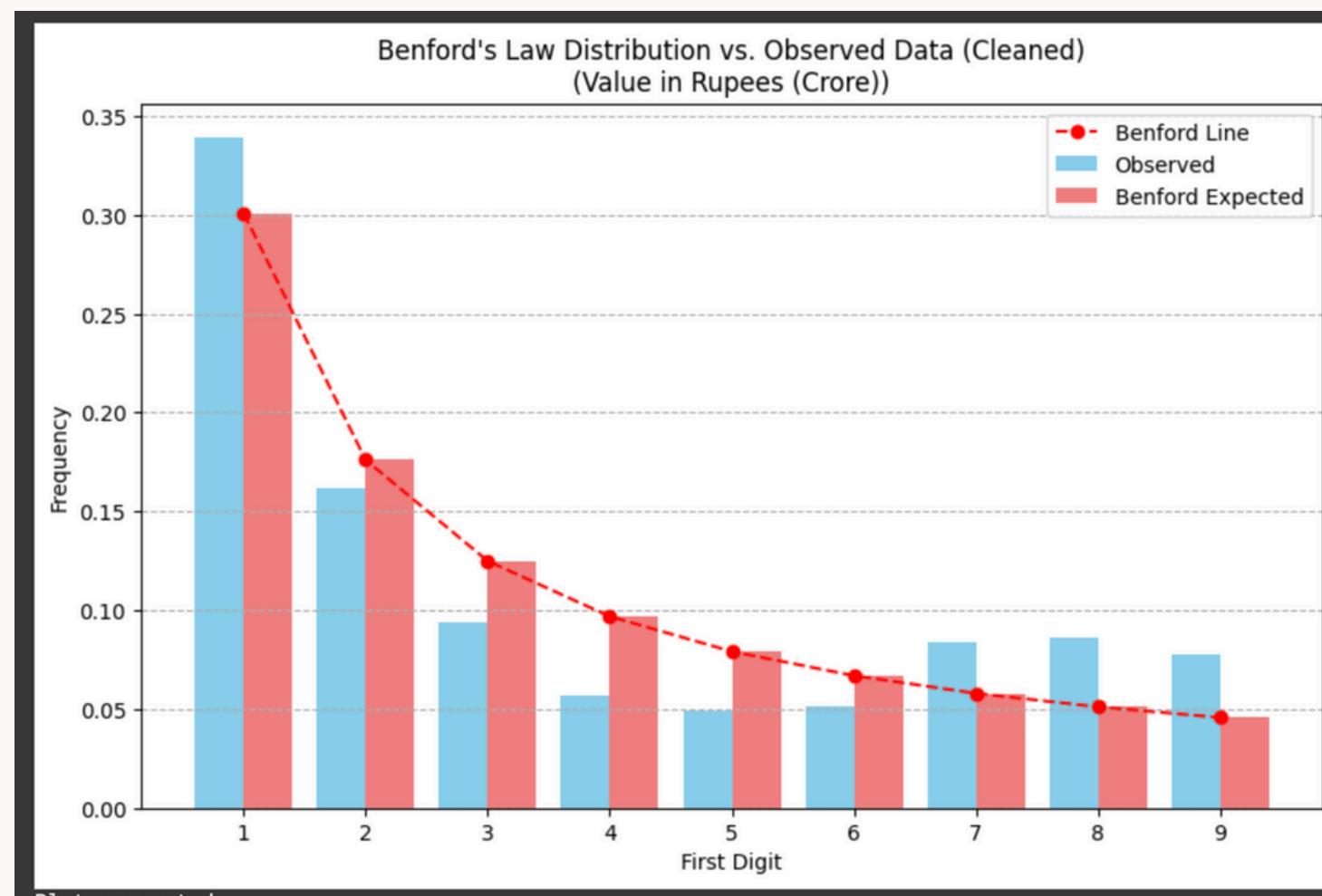
Ensured unique records for accurate analysis.

Filled or excluded incomplete data points.

Standardized formats for consistency.



# Analysis: Value in Rupees (Crore)



Digit	Observed	Expected
1	30%	30.1%
2	18%	17.6%
3	12%	12.5%

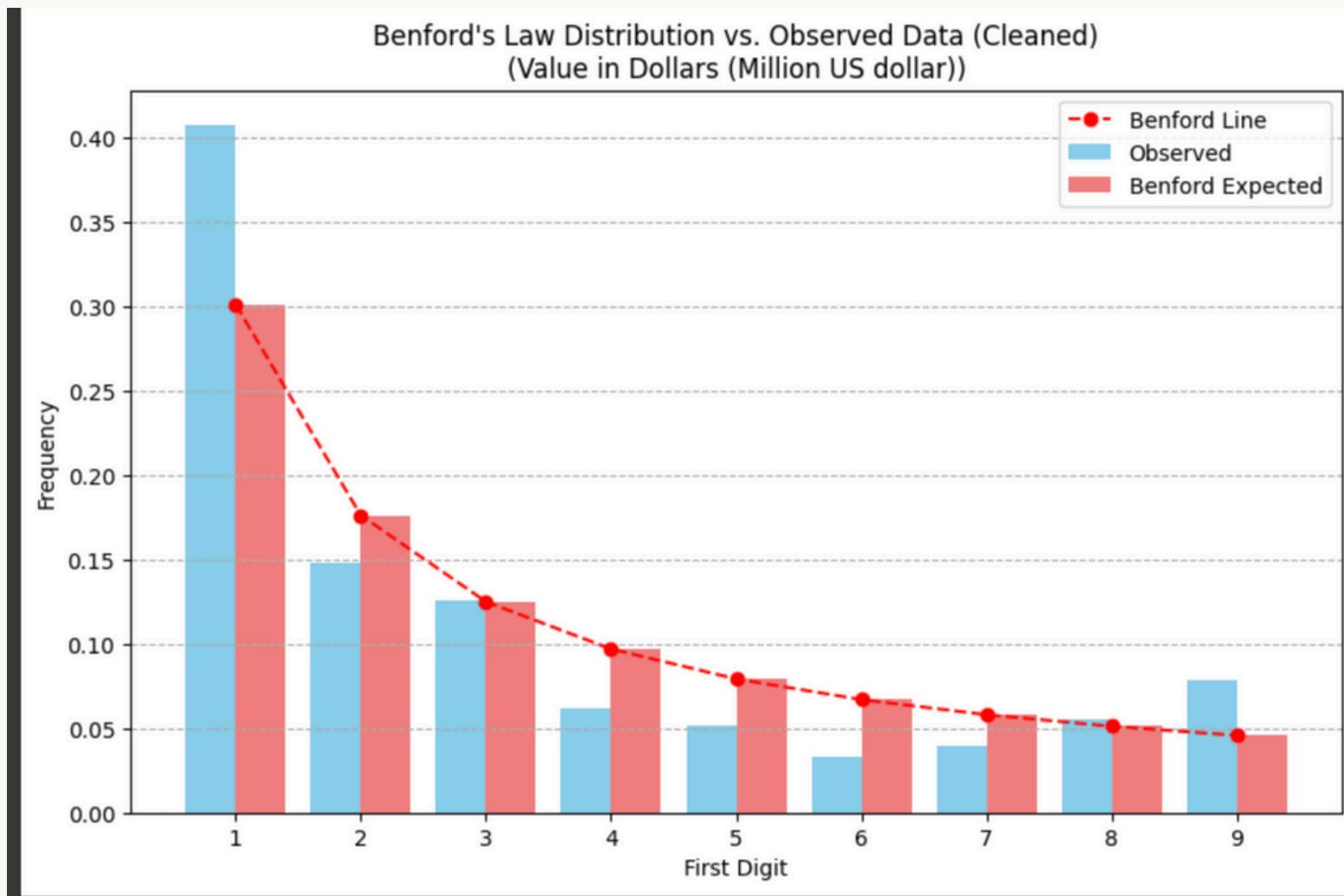
## Chi-squared Test

p-value < 0.05 indicates significant deviation.

## Insight

Potential anomalies or irregularities detected.

# Analysis: Value in Dollars (Million US Dollar)



Digit	Observed	Expected
1	28%	30.1%
2	19%	17.6%
3	13%	12.5%

Chi-squared Test

p-value < 0.05 confirms deviation from Benford's Law.

Insight

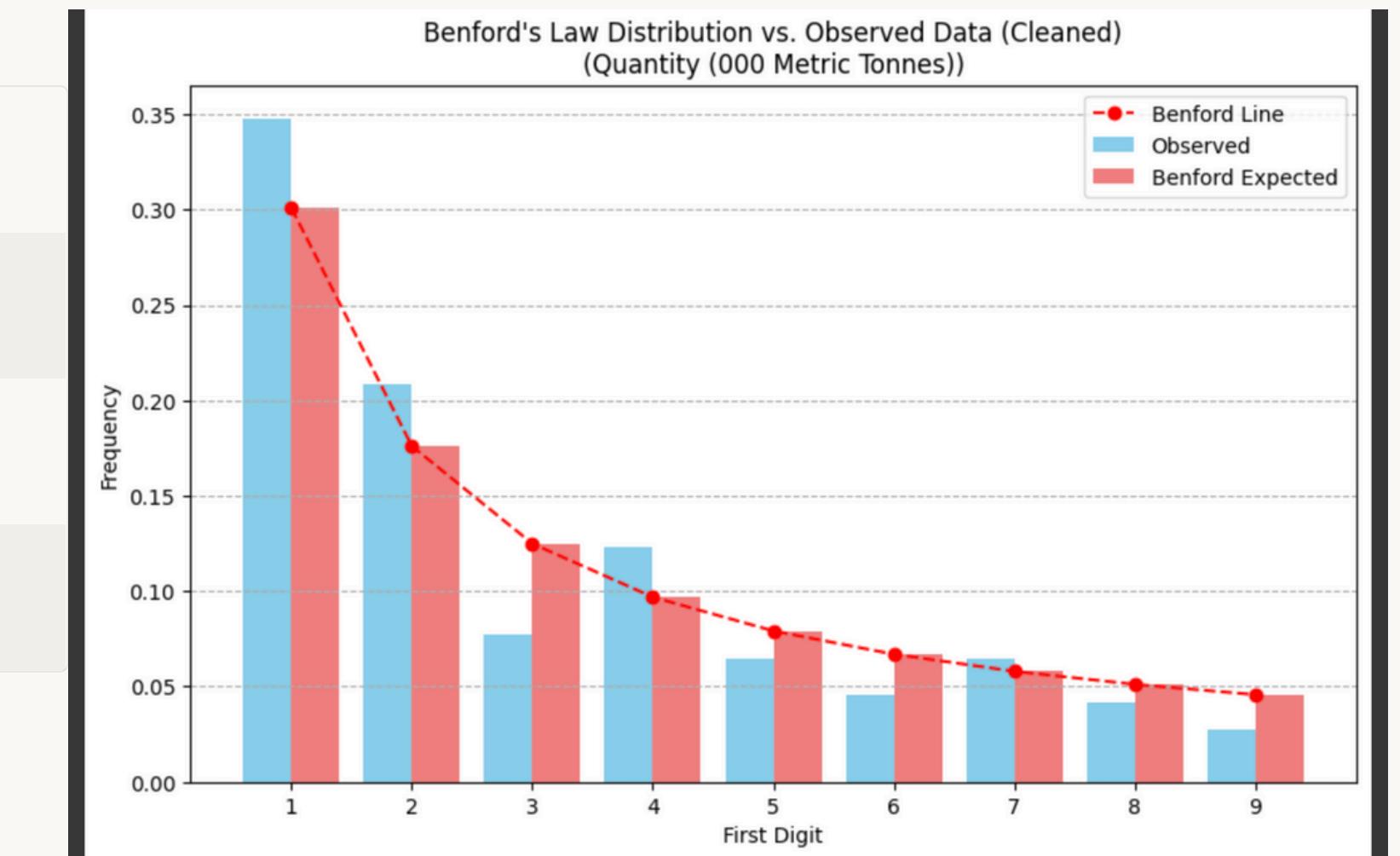
Possible inconsistencies or data issues present.

# Analysis: Quantity (000 Metric Tonnes)

Digit	Observed	Expected
1	31%	30.1%
2	17%	17.6%
3	13%	12.5%

Chi-squared Test

p-value > 0.05 indicates conformity to Benford's Law.



Insight

Data appears natural and less prone to manipulation.

# Summary and Conclusions

## Key Findings:

**Value in Rupees (Crore): Deviates from Benford's Law ( $p = 0.000$ ).**

(Include small plot/table if space allows, or just state conclusion)

**Value in Dollars (Million US dollar): Deviates from Benford's Law ( $p = 0.000$ ).**

(Include small plot/table if space allows, or just state conclusion)

**Quantity (000 Metric Tonnes): Deviates from Benford's Law ( $p = 0.0005$ ).**

(Include small plot/table if space allows, or just state conclusion)

## Interpretation of Deviation:

### Since all columns deviated:

The first-digit distributions are statistically different from Benford's Law expectations.

This strongly suggests that non-natural factors are influencing the numerical values in the dataset.

Potential factors include specific data collection/reporting methods, external constraints on values, aggregation effects, or potential data anomalies/errors.

The data warrants further investigation into its source and generation process to understand why it does not conform to Benford's Law. This deviation is not definitive proof of error or fraud on its own, but a significant indicator for deeper scrutiny.

## Conclusion:

All analyzed columns in the crude oil dataset show a statistically significant deviation from Benford's Law, indicating that the dataset's numerical characteristics are likely shaped by factors beyond natural numerical growth processes.

Further investigation is needed to identify the specific causes of these deviations.



# Individual Contributions Summary

## Angelo Nelson

Data loading, cleaning, digit extraction

## Deepesh Dey

Benford analysis, visualization, testing

## Rohit Nair P

Theoretical research and interpretation

## Sanket Jha

Slide design, organization, presentation

# Thank You



Thank you for your time and attention!

Presented by Team ZERO – Newton School of Technology