Analyzing Real-World Data Using Benford's Law

Team Name: Chowck

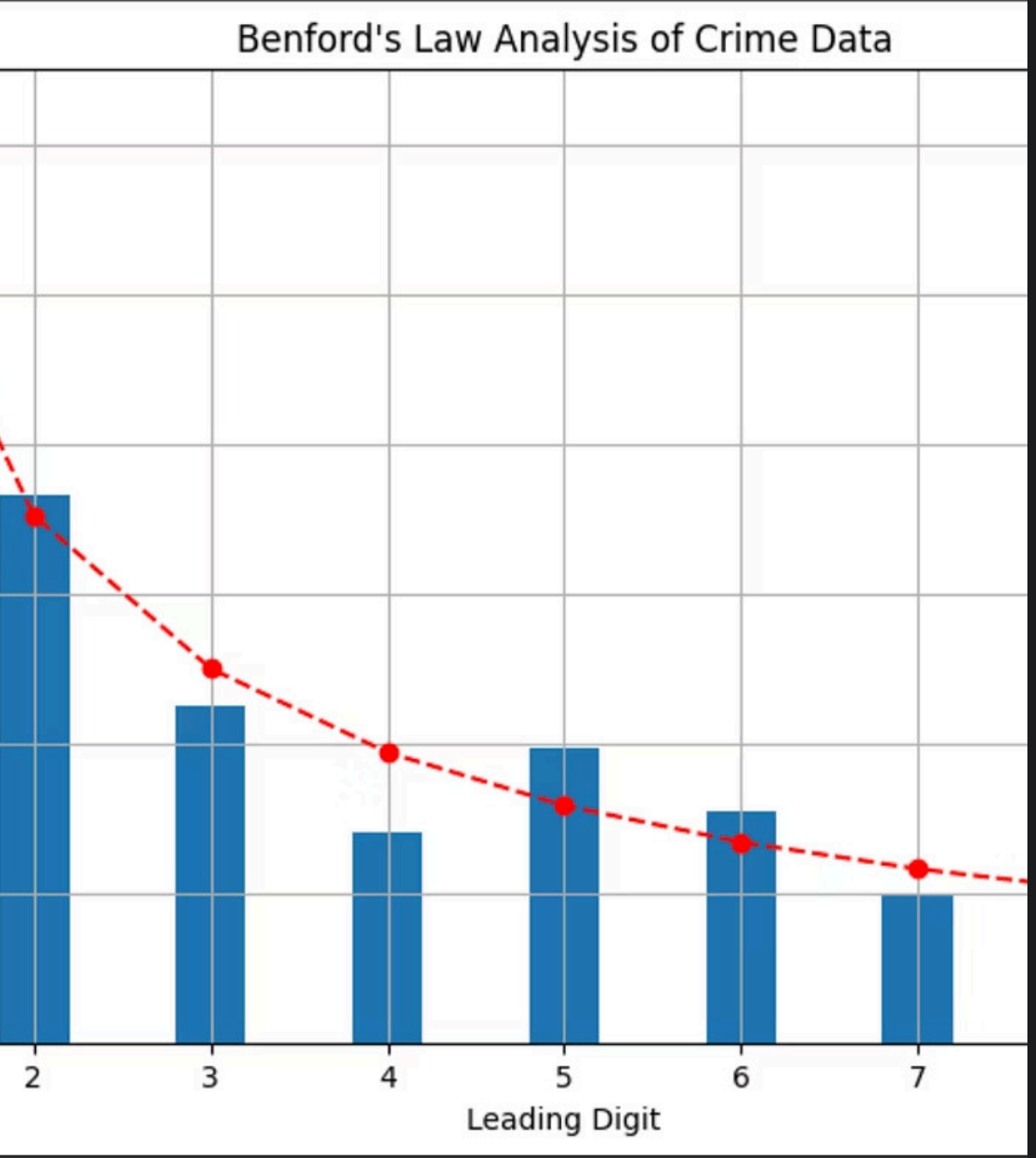
Team Members:

- Nilesh Nand Lal - nilesh.nandlal@adypu.edu.in
- Ankit Kumar - ankit.kumar01@adypu.edu.in
- Ankit Singh - ankit.singh01@adypu.edu.in
- Aditya Poddar - aditya.poddar@adypu.edu.in

Introduction:

Benford's Law:

Benford's Law, also known as the First-Digit Law, is a mathematical principle that describes how often certain digits appear as the first digit in many real-world datasets. According to this law, numbers that occur naturally in various contexts—such as financial records, population figures, scientific data, or physical measurements—tend to begin with smaller digits. Specifically, the number 1 appears as the leading digit about 30% of the time, while larger digits occur less frequently, with the number 9 appearing first in less than 5% of cases. This distribution follows a logarithmic formula: □ (□) = log 10 (1 + 1 □) P(d)=log 10 (1+ d 1), where □ d is the leading digit. Benford's Law is counterintuitive because one might expect all digits from 1 to 9 to appear with equal probability, but in practice, that's not the case. The law is particularly useful in fields like auditing, accounting, and forensic science, where it can help detect fraud or data manipulation—since fabricated numbers often deviate from this natural pattern. It applies best to datasets that span several orders of magnitude and are not artificially constrained.



Result

- **Digit 1** occurs slightly more often than predicted (~31% observed vs. 30.1% expected), which aligns well with Benford's Law.
- **Digits 2 and 3** also follow the expected trend closely, though digit 2 is slightly overrepresented.
- **Digits 4 to 9** mostly stay within a reasonable range compared to the expected values, though there are some minor deviations (e.g., digit 5 appears slightly more often than expected, digit 4 slightly less).
- **Overall trend**: The observed data generally follows the downward slope expected from Benford's Law, suggesting that the dataset does **not show major irregularities or signs of manipulation**.

Individual Contributions

Nilesh Nand Lal	• Select datasets suitable for Benford's Law (e.g., financial records, population data, election results). • Clean the data (remove missing values, ensure numerical format). • Extract the leading digits for analysis.
Ankit Kumar	• Made a full presentation and wrote every team member contribution and Analysis of data of Benford's Law
Ankit Singh	• Interpret whether the data follows Benford's Law using statistical tests (e.g., chi-square test). • Explain reasons for conformity or deviation. • Write the project report.
Aditya Poddar	• Calculate the frequency distribution of the first digits (1–9). • Compare observed distribution with expected Benford distribution. • Create graphs (bar plots or histograms) to visualize differences.

Thank You!