# COVID-19 Data Analysis: Benford's Law Validation

- **A** data analysis project exploring whether COVID-19 statistics follow Benford's Law distribution patterns.

**Team: Mohit Kourav, Kumar Gautam, Kumar Manak, Rohit Dahiya**

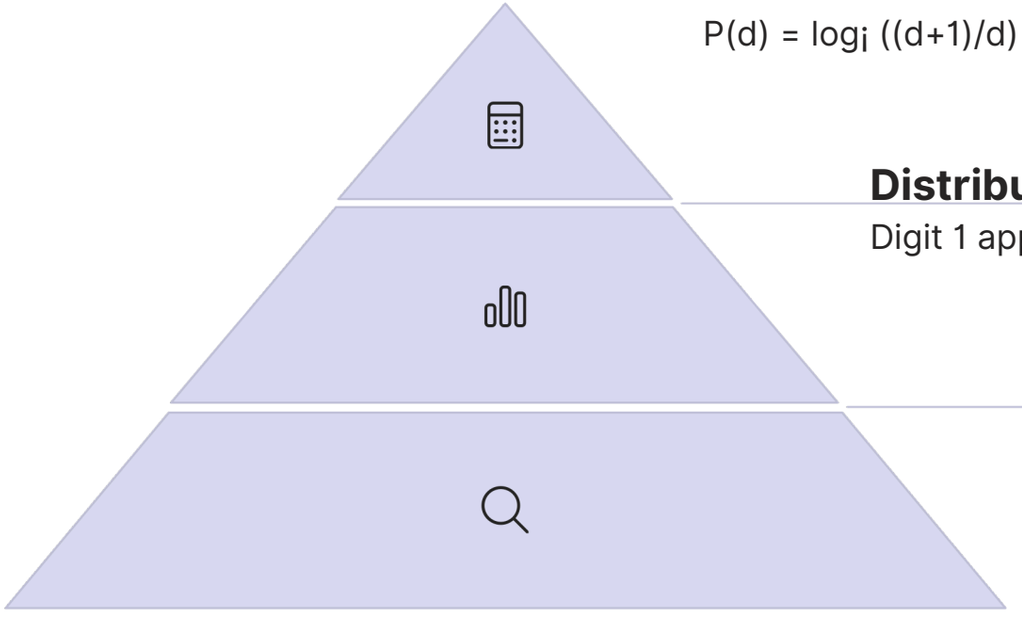
# Project Overview

### Objective
Analyze real-world COVID-19 data to check if it follows Benford's Law.

### Goal
Explore data anomalies and validate patterns using Benford's Law.

### Data Source
COVID-19 dataset including total cases, active cases, deaths, recoveries.

# Understanding Benford's Law



**Formula**

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

**Distribution**

Digit 1 appears ~30%, larger digits less frequently

**Applic ation**

Validates natural data patterns

Benford's Law states that in many datasets, the leading digit follows a predictable distribution. We'll check if COVID-19 data conforms to this pattern.

# Validation Methodology

### Extract First Digits

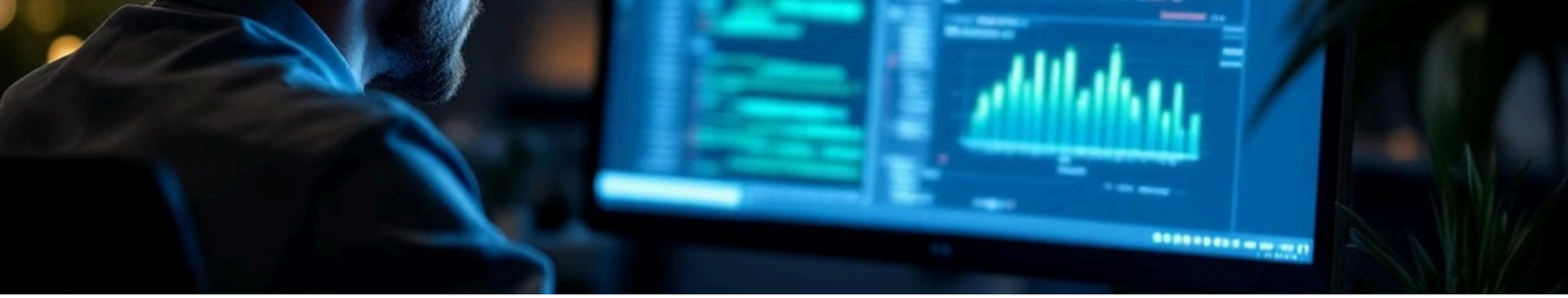
Isolate the first non-zero digit from each COVID-19 data point.

### Calculate Frequencies

Determine how often each digit (1-9) appears as a first digit.

### Compare Distributions

Contrast observed frequencies against Benford's Law expectations.



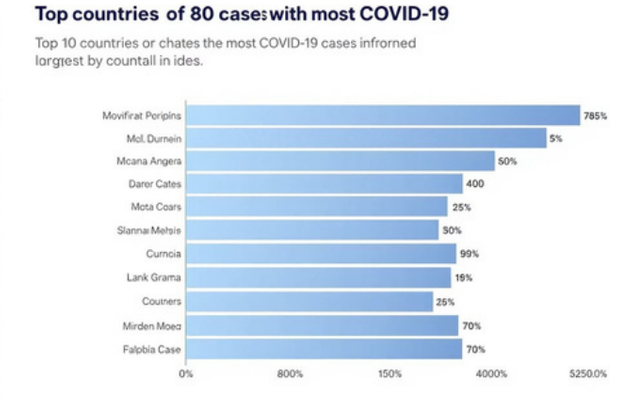
# Exploratory Data Analysis

### Data Preprocessing
Handled missing data and cleaned columns for analysis.

### Data Transformation
Applied necessary transformations to prepare for Benford analysis.

### Pattern Discovery
Identified initial trends in case distributions and growth patterns.

# Key Visualizations

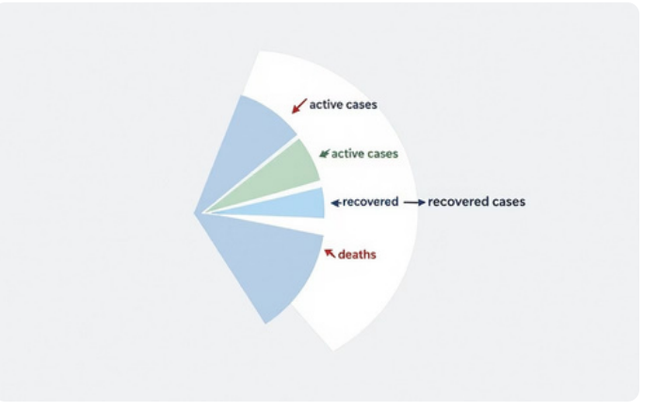



## Country Distribution

Cases varied significantly across countries, with clear regional patterns.


## Temporal Trends

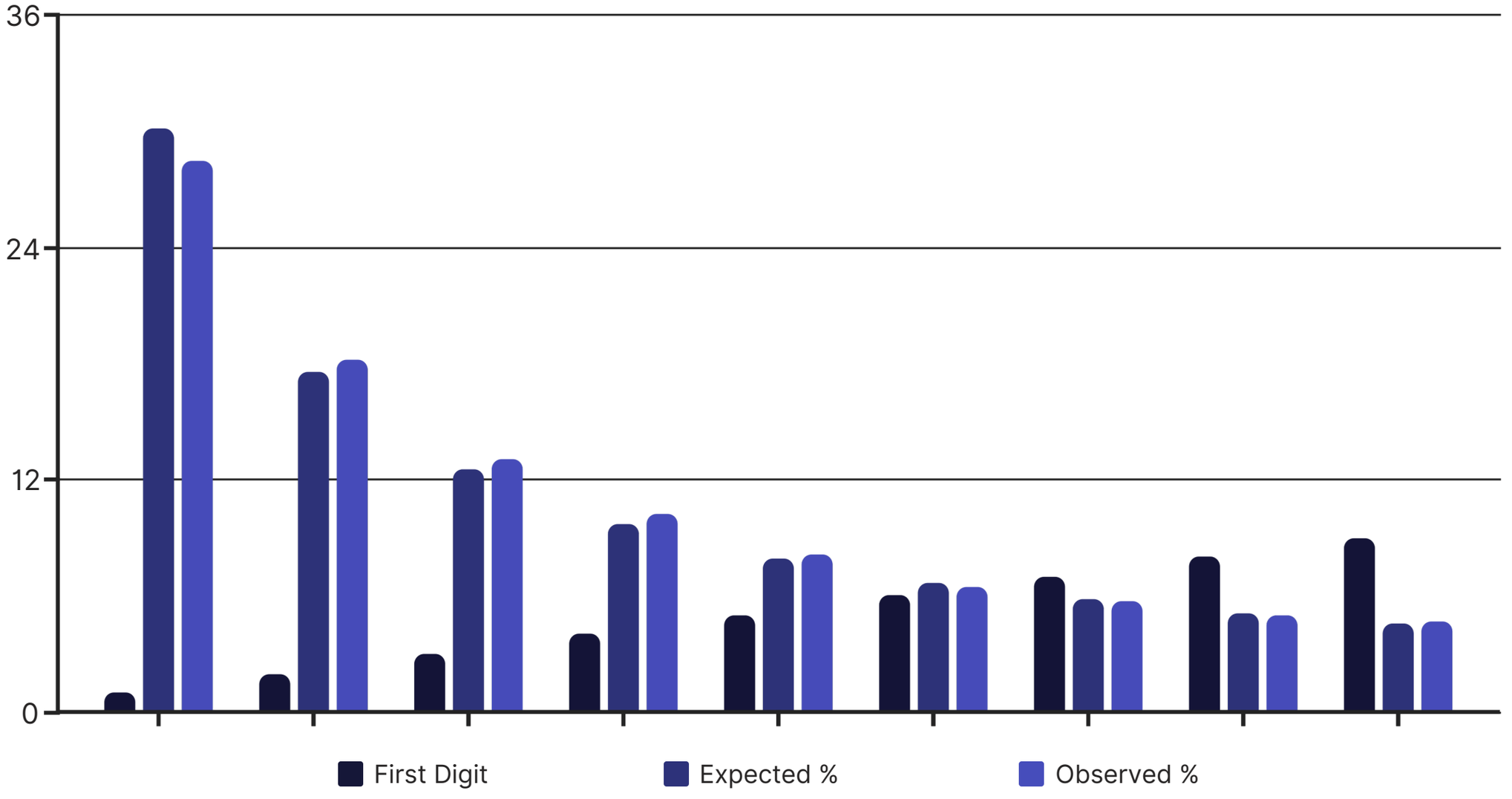
Case growth showed distinct waves across the pandemic timeline.




## Case Breakdown

Proportional analysis of active cases, recoveries, and fatalities.

# Benford's Law vs. Observed Data




The chart compares expected Benford's Law distribution with observed frequencies in COVID-19 data. Minor deviations exist but overall patterns show general conformity.

# Insights and Findings

### General Conformity

COVID-19 data largely follows Benford's Law, suggesting natural growth patterns in the pandemic.

### Minor Deviations

Small variations may indicate reporting inconsistencies or data collection challenges across regions.

### Validation Tool

Benford's analysis proves useful for identifying potential anomalies in pandemic reporting.

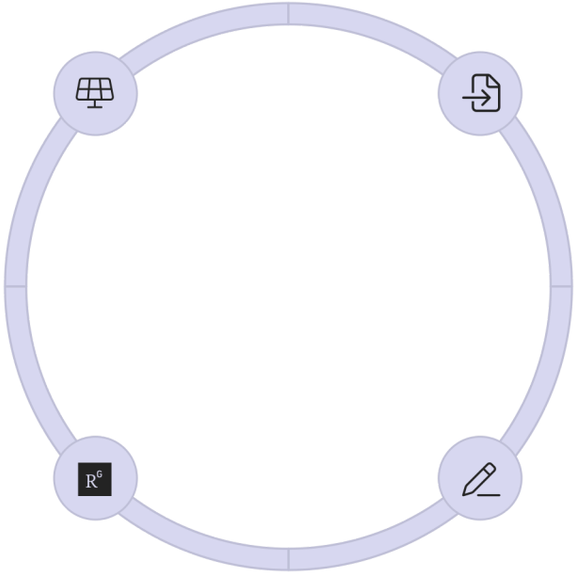
# Team Contributions

### Mohit Kourav

Led data cleaning, preprocessing, Benford's Law analysis, and visualization creation.

### Rohit Dahiya

Contributed to research, wrote conclusions and summary findings.



### Kumar Gautam

Assisted with data import, preprocessing, visualization, and results interpretation.

### Kumar Manak

Supported preprocessing, wrote insights and interpretation sections.

# Thank You!

We sincerely appreciate your time and attention. Please feel free to reach out with any questions or feedback you may have.