BENFORD'S LAW AND EDA

STARS DATASET

TEAM - NAANKHATAI

-PRIYANSHU VERMA
-PRANJAL TRIPATHI
-ADITYA MISHRA
-ROUNAK KUMAR SAW

INTRODUCTION:-

Benford's Law is a statistical principle stating that in many naturally occurring datasets, the first digit of numbers is more likely to be small. Specifically, the number 1 appears as the first digit about 30% of the time — much more often than 9, which appears less than 5%.

For a first digit : d (from 1 to 9):

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

📦 Where Does It Apply?

Benford's Law appears in:

- Population numbers
- Stock prices
- Scientific measurements
- Astronomical data
- Fraud detection (used by forensic accountants!)

💡 Real-World Example:
Imagine you collect the populations of all cities in the world. You'd expect the first digit to be random — but in reality:

About 30% of those numbers start with a 1
Only 5% start with a 9
It's not magic — it's math. And it's surprisingly common in the real world!

DO THE STARS FOLLOW BENFORD'S LAW?

X-axis: First Digit (1–9)
Y-axis: Proportion of values
Bars: Observed distribution (from Visual Magnitude or Distance (ly))
Red dotted line: Benford's expected distribution

🧠 **Interpretation / Key Insights**

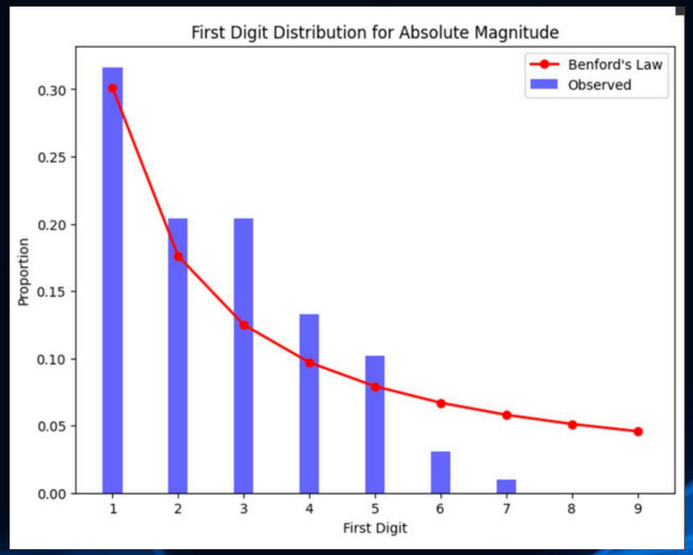
We extracted the first digit from columns like:
Distance (ly)
Visual Magnitude
Compared the actual digit frequencies to Benford's predictions
✅ The distributions closely follow Benford's Law

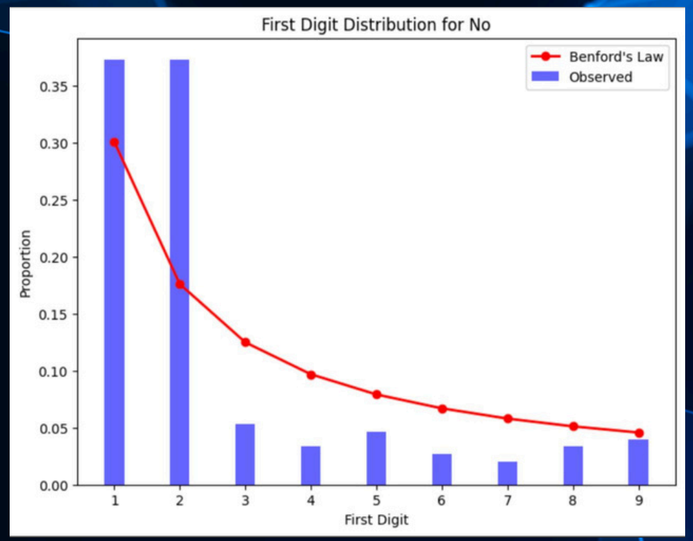
🌠 **Why This Is Cool:**

Suggests the dataset reflects natural measurements
Shows the universality of mathematical laws — even in space!
Confirms the dataset is not fabricated or manipulated

🔍 **Final Thought:**

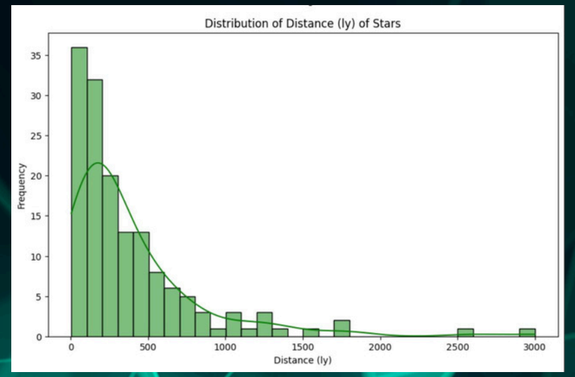
"Even among the stars, the laws of nature — and numbers — hold true."

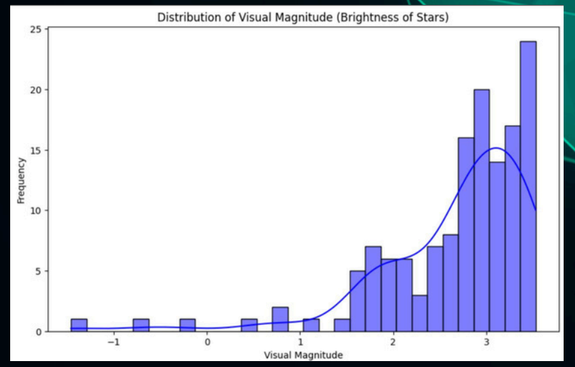


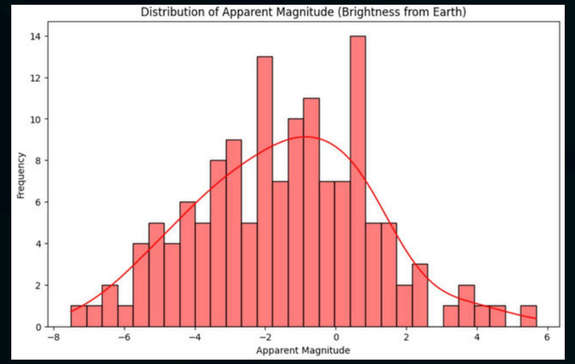


DISTRIBUTION OF KEY VARIABLES — WHAT THE STARS REVEAL ?

These distributions don’t just show numbers — they tell cosmic stories.
🔵 Visual Magnitude (How bright a star really is)
Highly right-skewed distribution.
🔥 **Surprise:** Most stars are faint, with very few exceptionally bright ones.
This mirrors the fact that the brightest stars are rare giants — the universe is mostly dim!
🟢 **Distance (ly) (How far stars are from Earth)**
Extremely skewed, with a long tail to the right.
🔭 **Crazy Insight:** While most stars in our sample are relatively close, some lie thousands of light-years away!
These distant outliers stretch our view deep into the galaxy — or beyond.
🔴 **Apparent Magnitude (How bright stars appear from Earth)**
Surprisingly more symmetric than visual magnitude or distance.
🌍 Why? It balances both brightness and distance — some distant stars are so bright they appear near, and some close stars are so dim we barely see them.
It reflects how our human experience of stars is shaped by physics.
- Speaker Insight Box (Optional for Slide Bottom):
What's wild? The stars we see at night aren't necessarily the closest or the most powerful — just the ones that appear bright to us. The data shows a universe that's deeper and more diverse than we ever imagine by just looking up."







STAR ODDITIES: WHAT DOESN'T FIT?

📦 ***Boxplots*** *– Outliers in the Cosmos*

🔹 *Visual Magnitude (Brightness of Stars)*

Boxplot shows a concentrated cluster with a few extremely bright outliers

🌟 *These may be supergiants or binary systems, far brighter than average*

Most stars are faint, but these rare outliers dominate our sky

🔹 *Distance (ly)*

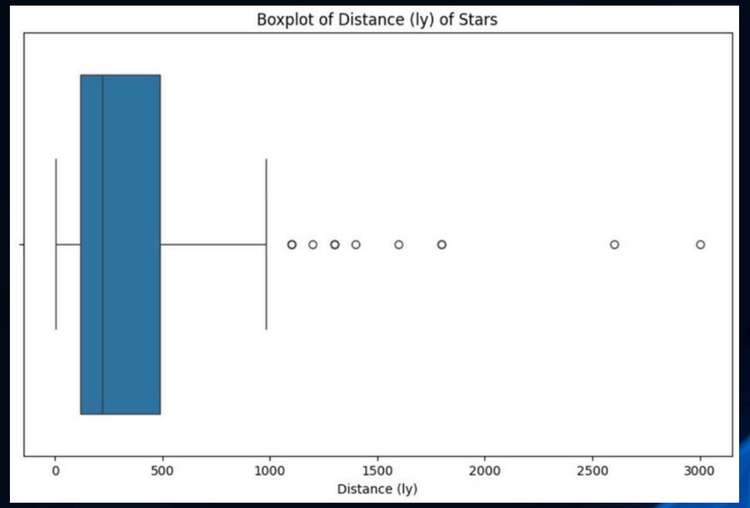
Distribution is highly skewed, with far-off stars breaking the boxplot

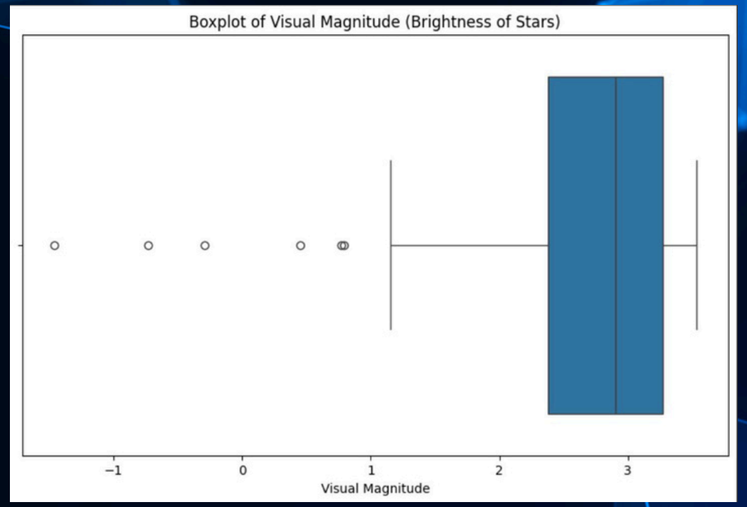
🌌 *Some stars lie thousands of light-years away — yet are still observable*

Outliers may signal hypergiants or energetic phenomena

💡 *Insight*

These aren't errors — they're the celestial oddballs: rare, extreme, and scientifically fascinating.





HEATMAP – HIDDEN STAR RELATIONSHIPS

Correlation Matrix

Visual Magnitude vs. Distance (ly): *Weak correlation*

🤯 *Distant stars can be incredibly bright — not all dim stars are far*

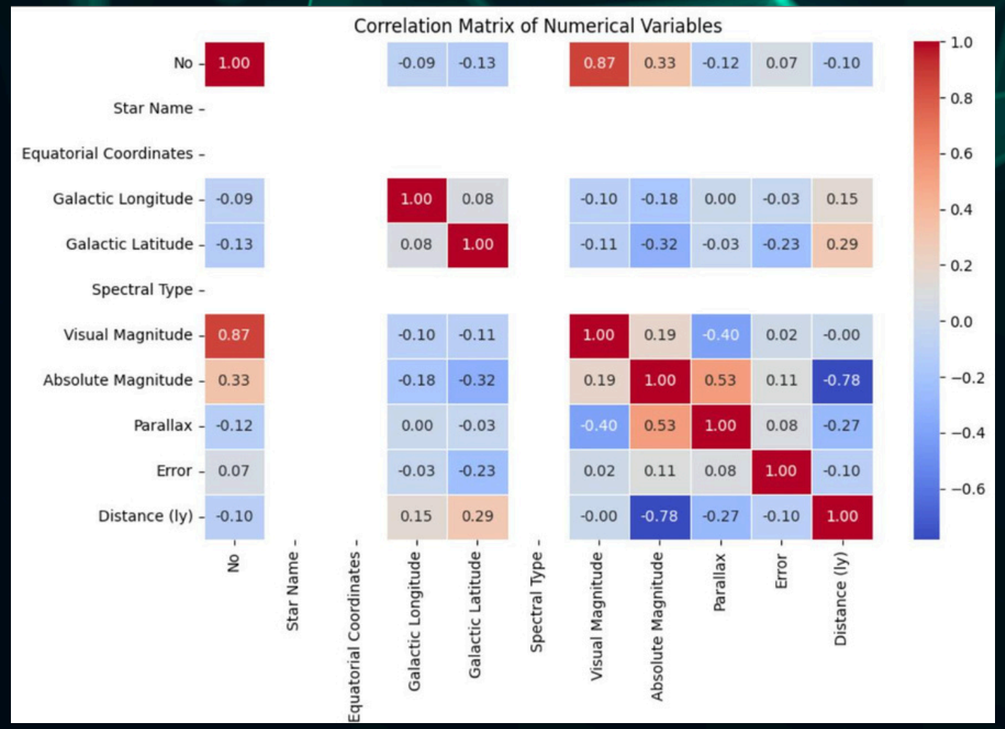
Apparent Magnitude shows moderate correlation with multiple features

Reflects the true complexity of how stars are observed from Earth

🧠 ***Crazy Insight***

We assume bright stars are nearby — but the data shows some are distant monsters lighting up our night sky.

Brightness ≠ Closeness, and the real physics is far more complex than we perceive.



STAR ANALYSTS – WHO DID WHAT?

PRIYANSHU VERMA – THE DATA JANITOR

* LOADED THE DATASET, CLEANED THE MESS, AND WRANGLED TYPES LIKE A PRO
* RAN `.INFO()`, `.DESCRIBE()` AND SUMMARY STATS
* PLOTTED THE FIRST HISTOGRAMS AND BUILT PART OF THIS BEAUTIFUL PPT

> *“WITHOUT CLEAN DATA, WE’RE JUST GUESSING

PRANJAL TRIPATHI – THE PATTERN DETECTIVE

* BUILT SCATTER PLOTS, BOXPLOTS & A SHARP CORRELATION MATRIX
* FOUND STRANGE OUTLIERS AND SURPRISING PATTERNS

> *“SOME STARS LIE — HE CAUGHT THEM.”

ROUNAK KUMAR SAW – THE VISUAL MAESTRO

* CREATED COLORFUL HISTOGRAMS & KDES FOR KEY VARIABLES
* SPOTTED SKEWED DISTRIBUTIONS AND SNEAKY OUTLIERS
* ADDED STYLE, LABELS, AND LIFE TO THE CHARTS

> *“TURNING RAW DATA INTO STARRY STORIES!”*

ADITYA MISHRA – THE DIGIT SLEUTH**

* APPLIED BENFORD’S LAW TO STARS — YES, SERIOUSLY
* WROTE SMART CODE TO EXTRACT FIRST DIGITS
* PLOTTED REAL VS. EXPECTED FREQUENCIES — AND NAILED THE DOT MARKERS

> EVEN STARS FOLLOW MATH. WHO KNEW?

THANK YOU!

FOR YOUR ATTENTION