E0422037
Aaditya Balakrishnan

1. Why is Data Science Called the New Electricity?

Data Science is often called **"the new electricity"** because just as electricity transformed industries in the 20th century, data-driven intelligence is now powering the 21st-century economy. Electricity was once an innovation restricted to lighting homes, but soon it became a universal enabler for transportation, manufacturing, healthcare, and communication. Similarly, data science has evolved from being a niche tool for statisticians to a universal driver of progress across every sector today.

Timeline of Evolution

- **1960s–1980s:** Foundations of statistical analysis and database systems laid the groundwork. Early adoption in scientific research and government record-keeping.
- **1990s:** The internet boom led to exponential growth in digital data. Machine learning algorithms emerged as practical tools for pattern recognition.
- **2000s:** Big Data technologies (Hadoop, NoSQL) enabled storage and processing of massive datasets. Companies began using analytics for business intelligence.
- **2010s:** AI and deep learning matured, allowing predictive and prescriptive analytics at scale. Cloud computing made data science accessible worldwide.
- **2020s and beyond:** Integration of IoT, real-time analytics, and generative AI marks the era of data science as a utility, similar to electricity — invisible yet indispensable.

Real-World Applications

1. **Healthcare:** Predictive analytics helps in early detection of diseases, such as using machine learning on ECG and imaging data to assess cardiovascular risk. Personalized medicine tailors treatments to patient profiles, reducing trial-and-error in therapies.
2. **Finance:** Fraud detection systems continuously analyze millions of transactions in real-time, flagging suspicious activities. Algorithmic trading leverages data-driven predictions to maximize profit and minimize risk.
3. **Marketing & Government:** Retailers like Amazon and Netflix use recommendation engines to personalize user experiences, boosting customer engagement. Governments apply data science in areas like smart city management, resource allocation, and even monitoring public health trends (e.g., during COVID-19).

Conclusion

Just as electricity became the backbone of industrial progress, data science is now the invisible infrastructure driving innovation across healthcare, finance, marketing, and governance. Its evolution over the past decades shows how it has shifted from an experimental science to a universal enabler — making it truly "the new electricity."

2.

E0422037
Aaditya Balakrishnan

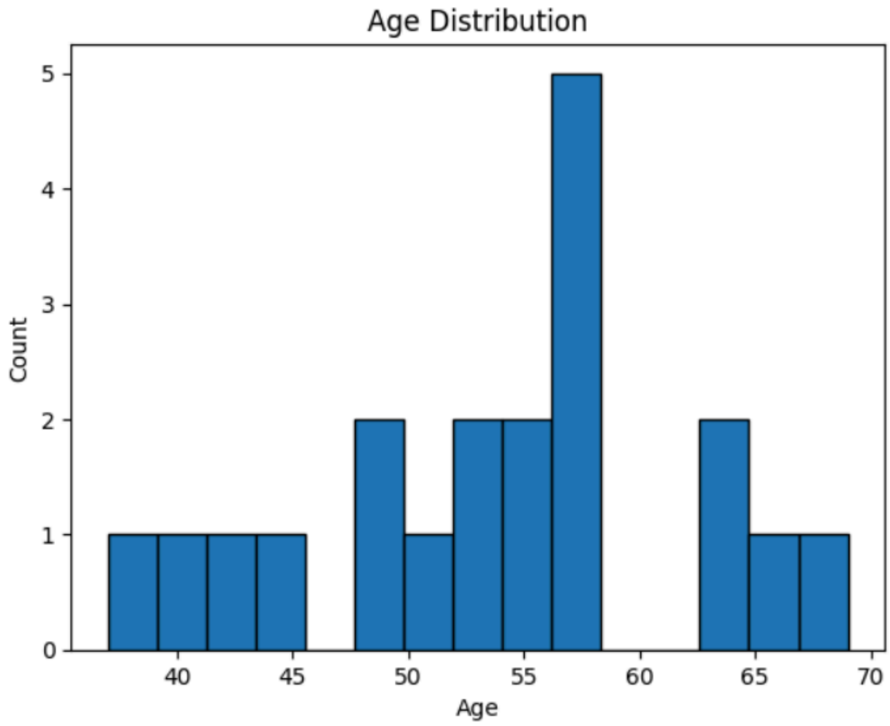
E0422037 Aaditya Balakrishnan

```
from google.colab import files
uploaded = files.upload()
```

Choose Files heart.csv
heart.csv(text/csv) - 820 bytes, last modified: n/a - 100% done
Saving heart.csv to heart.csv

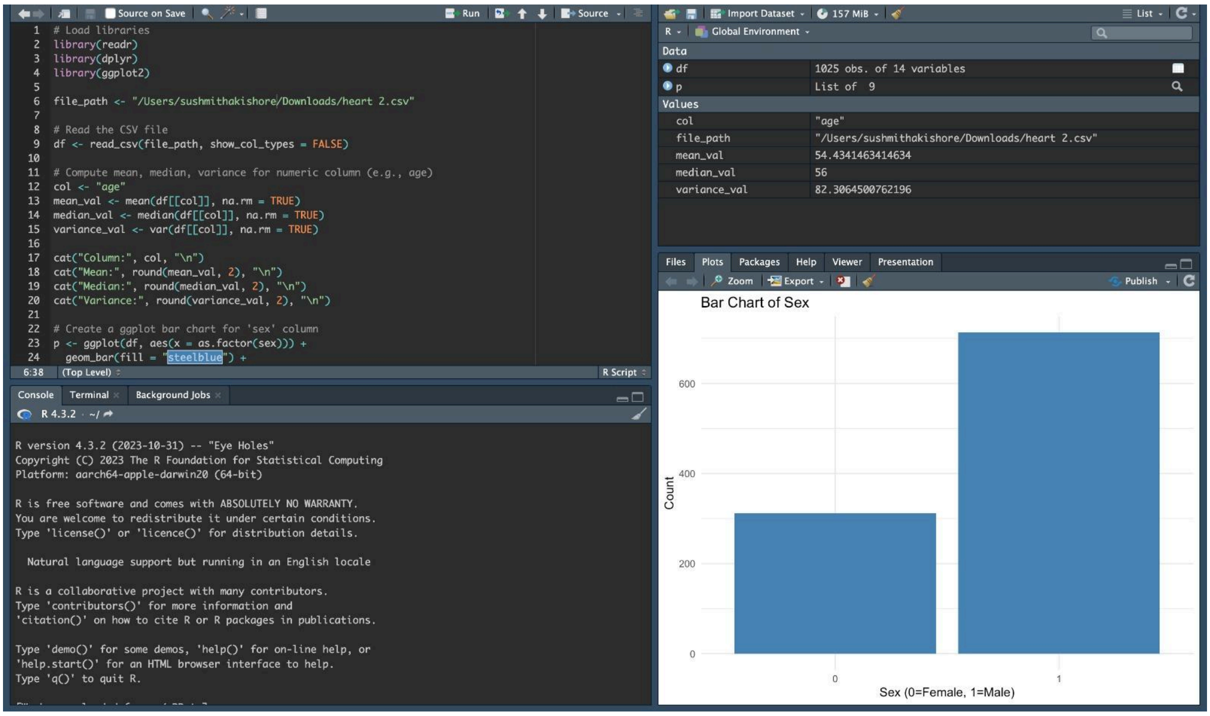
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


df = pd.read_csv("heart.csv")


print("Missing values in each column:")
print(df.isnull().sum())


corr_matrix = df.corr()
top3_corr = corr_matrix['target'].abs().sort_values(ascending=False)[1:4]  # exc
print("\nTop 3 correlations with target:")
print(top3_corr)


plt.hist(df['age'], bins=15, edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```

```
Missing values in each column:
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64

Top 3 correlations with target:
age        0.349332
fbs        0.342997
thalach    0.341990
Name: target, dtype: float64
```



R code
# Load libraries
library(readr)

```r
library(dplyr)
library(ggplot2)

file_path <- "/Users/AadityaBalakrishnan/Downloads/heart 2.csv"

# Read the CSV file
df <- read_csv(file_path, show_col_types = FALSE)

# Compute mean, median, variance for numeric column (e.g., age)
col <- "age"
mean_val <- mean(df[[col]], na.rm = TRUE)
median_val <- median(df[[col]], na.rm = TRUE)
variance_val <- var(df[[col]], na.rm = TRUE)

cat("Column:", col, "\n")
cat("Mean:", round(mean_val, 2), "\n")
cat("Median:", round(median_val, 2), "\n")
cat("Variance:", round(variance_val, 2), "\n")

# Create a ggplot bar chart for 'sex' column
p <- ggplot(df, aes(x = as.factor(sex))) +
  geom_bar(fill = "steelblue") +
  labs(title = "Bar Chart of Sex", x = "Sex (0=Female, 1=Male)", y = "Count") +
  theme_minimal()

print(p)

# Save the plot (optional)
ggsave("/Users/sushmithakishore/Downloads/bar_chart.png", plot = p, width = 6, height = 4,
dpi = 150)
```



3. Role Exploration: Data Scientist

A **Data Scientist** is a professional who transforms raw data into meaningful insights that drive decision-making. Much like electricity powered the industrial age, data science now powers the digital era, enabling organizations to predict trends, optimize operations, and create intelligent systems.

Responsibilities

The core responsibilities of a Data Scientist include collecting and cleaning large datasets, performing exploratory data analysis (EDA), and applying statistical and machine learning models to uncover hidden patterns. They design predictive systems, build pipelines for data processing, and ensure reproducibility of results. Another key responsibility is communicating findings through compelling dashboards, visualizations, and reports so that non-technical stakeholders can act on the insights. Data Scientists also collaborate with engineers, analysts, and business leaders to align data-driven strategies with organizational goals.

Skills & Tools

Data Scientists require strong programming skills in **Python, R, and SQL** to manipulate data and implement models. They must be proficient in **machine learning frameworks** like Scikit-learn, TensorFlow, or PyTorch, and capable of building end-to-end solutions. Knowledge of **data visualization** tools such as Tableau, Power BI, or matplotlib is essential for presenting results. Equally important are core skills in **statistics, probability, linear algebra, and data wrangling**. On the business side, strong problem-solving abilities and communication skills are critical for turning insights into actionable strategies. Familiarity with cloud platforms such as AWS, Azure, or GCP further strengthens their ability to scale solutions.

Real LinkedIn Job Post (Illustrative Example)

Job Title: Data Scientist – Global Tech Company

Highlighted Requirements:

1. Proficiency in Python and SQL for building predictive models.
2. Experience with machine learning techniques (classification, regression).
3. Strong data visualization skills using Tableau or matplotlib.
4. Ability to communicate complex results to non-technical stakeholders.
5. Knowledge of cloud-based tools and scalable data pipelines.