

1. List ten uses of data science for industry.

- I. Manufacturing (Predictive Maintenance): Data science can help optimise supply chains, predict demand, and assess production efficiency. Data scientists can also collect data to calculate machine efficiency and analyse defect levels. Data science models predict equipment failures before they happen by analysing sensor data, reducing downtime and maintenance costs.
- II. Finance (Fraud Detection): Data science can help improve credit scoring, detect fraud, and provide more efficient solutions to end-users. Data science algorithms detect unusual transaction patterns in real-time, helping prevent fraud and ensuring the security of financial systems.
- III. Supply chain management: Data science can help identify bottlenecks and inefficiencies in supply chains by analysing data about supplier performance and demand.
- IV. Logistics: Data science can help optimise operations by determining the most efficient delivery routes, managing fuel usage, and making more accurate predictions about supply and demand. Machine learning models optimise inventory levels and shipping routes, reducing costs and improving efficiency in supply chain management.
- V. Retail/Marketing (Customer Segmentation): By analysing customer data, businesses can segment their customer base for targeted marketing, improving personalisation and sales conversion. Data science can help enhance customer analytics and personalise marketing strategies.
- VI. Energy Consumption Forecasting (Utilities): Data science models predict energy consumption patterns, helping utility companies optimise energy production and distribution to meet demand.
- VII. Risk Assessment (Insurance): Data science models assess the probability of insurance claims or accidents based on customer data, helping insurance companies set premiums and manage risks.
- VIII. E-commerce (Product Recommendation Systems): By analysing customer preferences and behaviour, data science helps create personalised product recommendations, boosting sales and customer engagement.
- IX. Healthcare (Disease Diagnosis and Treatment): Machine learning models analyse medical records and imaging data to assist doctors in diagnosing diseases and suggesting personalised treatment plans. Data science can help enable predictive analytics for better patient care and resource allocation. Data scientists can efficiently speed up the processing of clinical and lab reports and utilise deep machine learning to yield additional diagnosis insights.
- X. Media and Entertainment (Sentiment Analysis): Data science helps businesses analyse customer feedback, social media posts, and reviews to understand public sentiment and trends.
- XI. Aerospace (Predictive Maintenance): About one-third of flight delays are due to unscheduled maintenance but data from IOT sensors can be used to better predict where trouble is likely to occur. Data science applications deliver

information to technicians in real-time and send system notifications to ensure the right tools and parts are available for the work.

- XII. Travel and transportation: (Estimating Travel Time): Many mobile devices include map applications that use data science models to estimate travel time and suggest not only various routes but also travel time options for air travel, driving, walking, or public transit.

2. Identify ten dataset from Kaggle and their applications.

I. AI-Powered Job Market Insights:

<https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights>

- Job Market Trends Analysis: Identifying trends in the job market, such as demand for specific skills, emerging industries, and job opportunities.
- Salary Prediction: Developing models to predict salaries based on job roles, locations, and experience levels.
- Career Path Recommendations: Using AI to suggest career transitions or development opportunities based on past job market data.
- Job Matching Algorithms: Enhancing recruitment platforms by creating smarter job matching systems between candidates and employers.

II. Huge Stock Market Dataset:

<https://www.kaggle.com/datasets/borismarjanovic/price-volume-data-for-all-us-stocks-etfs>

- Stock Price Prediction: Creating models to forecast future stock prices based on historical data.
- Portfolio Optimisation: Using the dataset to suggest optimal investment strategies by diversifying portfolios across industries or stocks.
- Market Trend Analysis: Analysing patterns such as market crashes, bull/bear markets, and overall market trends.
- Algorithmic Trading: Developing automated trading algorithms that can react to market movements and execute trades efficiently.

III. Summer Olympics Medals (1896-2024):

<https://www.kaggle.com/datasets/stefanydeoliveira/summer-olympics-medals-1896-2024>

- Historical Trends in Olympic Performance: Identifying countries' long-term performance trends and dominance in specific sports.
- Sports Popularity Analysis: Assessing how the popularity and competitiveness of various sports have changed over the years.
- Country Performance Prediction: Using past data to predict which countries might perform well in future Olympics.
- Impact of Socioeconomic Factors on Olympic Success: Studying the relationship between a country's economy, population, and Olympic success over time.

IV. Crimes Against Women in India (2001-2021):

<https://www.kaggle.com/datasets/balajivaraprasad/crimes-against-women-in-india-2001-2021>

- Crime Rate Analysis: Investigating the trends in crimes against women over two decades and identifying areas with rising or falling crime rates.
 - Policy Impact Evaluation: Analysing how changes in laws and policies affected the rates of crimes against women in India.
 - Predictive Policing: Building models to predict potential crime hotspots or periods when crimes against women might surge.
 - Awareness Campaign Design: Helping design more effective public awareness campaigns by understanding patterns of crimes against women and targeting regions or demographics accordingly.
- V. New York City Taxi Trip Duration:
<https://www.kaggle.com/c/nyc-taxi-trip-duration>
- Predicting taxi trip durations using regression models can help optimise transportation networks, improve taxi dispatch systems, and predict demand in ride-hailing services like Uber or Lyft. It's useful for traffic planning, route optimisation, and understanding patterns in urban mobility.
- VI. Amazon Product Review Data:
<https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews>
- This dataset is useful for sentiment analysis and product recommendation systems. Companies can analyse customer feedback to improve products, understand customer preferences, and make recommendations. It's also used for building natural language processing (NLP) models that can detect fake reviews or categorise product features.
- VII. HR Analytics: Employee Attrition & Performance:
<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
- This dataset helps organisations predict employee turnover using classification models. By analysing factors like job satisfaction, performance, salary, and work environment, companies can take preventive actions to reduce attrition, improve employee satisfaction, and better manage human resources.
- VIII. IMDB Movie Reviews Sentiment Analysis:
<https://www.kaggle.com/code/lakshmi25npathi/sentiment-analysis-of-imdb-movie-reviews>
- This dataset is valuable for sentiment analysis to gauge public opinion on movies. By classifying reviews as positive or negative, it helps streaming platforms, production houses, and marketers understand audience reception, tailor recommendations, and even predict box office success.
- IX. COVID-19 Global Dataset:
<https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset>
- Tracking and analysing the global spread of COVID-19, understanding patterns of infection, recovery, and fatalities. Governments and health organisations can use this data for modelling the spread of the virus, predicting future outbreaks, assessing the impact of lockdowns or vaccinations, and resource allocation (e.g., ventilators, vaccines, hospital beds). It can also be used for research purposes to understand the long-term impacts of COVID-19 globally.

3. **Research paper from 2024 on Data Science and AI with its important points.**

- Research paper: "Data Science and Artificial Intelligence in Scientific Research" by Tamara Magash

LINK-

https://www.researchgate.net/publication/380075282_Data_Science_and_Artificial_Intelligence_in_Scientific_Research_Introduction_to_The_Role_of_Data_Science_in_Scientific_Research

- **Role of Data Science in Research:** Data science has become essential in scientific research due to its ability to analyse large, complex datasets. It enhances traditional methods by allowing researchers to discover patterns and correlations that would have otherwise been missed. This has led to data-driven insights across various disciplines.
- **Predictive Modelling and Simulations:** Data science enables researchers to create predictive models and simulations that mimic real-world scenarios. These models are applied in fields such as weather forecasting, disease outbreak prediction, and manufacturing optimization. Simulations are also crucial for testing hypotheses in fields like physics and biology, helping researchers gain a deeper understanding of complex systems.
- **Data Mining and Knowledge Discovery:** Data mining refers to the process of discovering patterns and relationships in large datasets. Knowledge discovery goes further by extracting actionable insights from this data. These methods help researchers make sense of massive amounts of information, often revealing hidden trends that were previously unidentifiable through conventional analysis.
- **Collaborative Research and Open Science:** Data science has fostered a culture of collaboration among researchers, with large datasets being shared globally. Open science encourages the sharing of research methodologies and results, making research more transparent and reproducible. This promotes faster scientific advancements as researchers can build on each other's work.
- **Applications of AI in Research:** Artificial Intelligence (AI) has various applications in scientific research, such as drug discovery, environmental monitoring, and genomics. AI's ability to analyze massive datasets quickly and make predictions has accelerated discoveries, especially in fields where data collection is enormous, such as genomics and climate science.
- **Machine Learning in Research:** Machine learning techniques, including supervised and unsupervised learning, are widely used in scientific research. For example, in genomics, machine learning helps identify genetic variations linked to diseases. In environmental science, it helps predict climate changes by analyzing data from satellites and sensors.
- **AI in Drug Discovery:** AI has transformed drug discovery by predicting the efficacy and safety of chemical compounds, leading to faster and more cost-effective drug development. Machine learning algorithms can identify potential drug molecules by recognizing patterns in chemical structures and their biological activity.

- **Ethical Considerations in AI and Data Science:** As AI and data science become more integrated into scientific research, ethical issues such as data privacy, algorithmic bias, and transparency become increasingly important. Protecting personal data and ensuring fairness in AI algorithms are critical to prevent discrimination and ensure responsible use of these technologies.
- **Automation and Robotics:** AI-powered robots are becoming essential in scientific research for performing repetitive tasks, allowing researchers to focus on more complex experiments. For example, robots can automate DNA sequencing in genomics or test new materials in laboratories, accelerating discovery processes.