

# Introduction to Data Analytics

## T5 Bootcamp by SDAIA



**SDAIA**  
الهيئة السعودية للبيانات  
والذكاء الاصطناعي  
Saudi Data & AI Authority

Let's start together...



# T5

## The Big Picture

### Bootcamp layout



- Bootcamp duration: 12 weeks
- Week 11: Module-based learning with hands-on exercises, projects, tasks, exam.
- Week 12: Senior Capstone Project
- Focus on practical learning through presentations, quizzes, and projects
- Each module consists of:
  - PowerPoint presentations covering theoretical concepts
  - Interactive Jupyter notebooks for practical exercises
  - Assigned tasks to reinforce learning
  - Module project to apply knowledge in real-world scenarios
  - Module quiz to assess understanding
  - Presentation of module project to peers and instructors

### Module Environment

- Python Programming
- GoogleColab
- JupyterNotebook



# T5 Orientation

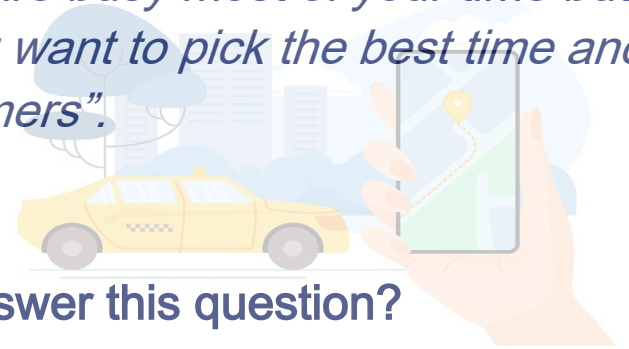
- Duration : 3 Months (12 Weeks)
- General Modules
  - Introduction to AI/ML
  - Introduction to Data Analytics
  - Introduction to Databases & SQL
  - Python Programming
  - Data Analysis using Python for ML
  - Machine Learning
  - Deep learning
  - Advanced Deep Learning
  - Generative AI and Computer Vision
  - NLP and Generative NLP
  - MLOPS
  - Senior Capstone
- Grading system
  - Attendance
  - Discord
  - Research
  - Lab
  - Communication and Interactions Skills
- Module Architecture
  - Slides
  - Hands-on Practice
  - Project
  - Task
  - Exam



# Thought Experiment

*“You work as a driver for a ride-sharing company. You are busy most of your time but can dedicate 5 hours per day to make your trips. You want to pick the best time and place to start accepting customers”.*

What is the approach that you will take to answer this question?



# Data is the New Black Gold!



<https://data.gov.sa/>



## National Data Bank

The National Data Bank (NDB) is a constellation of interconnected robust national data platforms that aim at accelerating the data literacy and instilling data as a common denominator for the digital economy in the Kingdom.



**50+**

Agencies Onboarded



**230+**

Datasets Hosted



**200+**

Marketplace APIs



**6K+**

Open Datasets



**130+**

Systems Cataloged



**2K+**

Standards Developed



**Open Data Portal**

Press here for more details



# What is Data Analytics?



# What is Data Analytics?

*The science of analysing raw data to extract insights and make conclusions, predictions, and support decisions.*

**Data → Value**

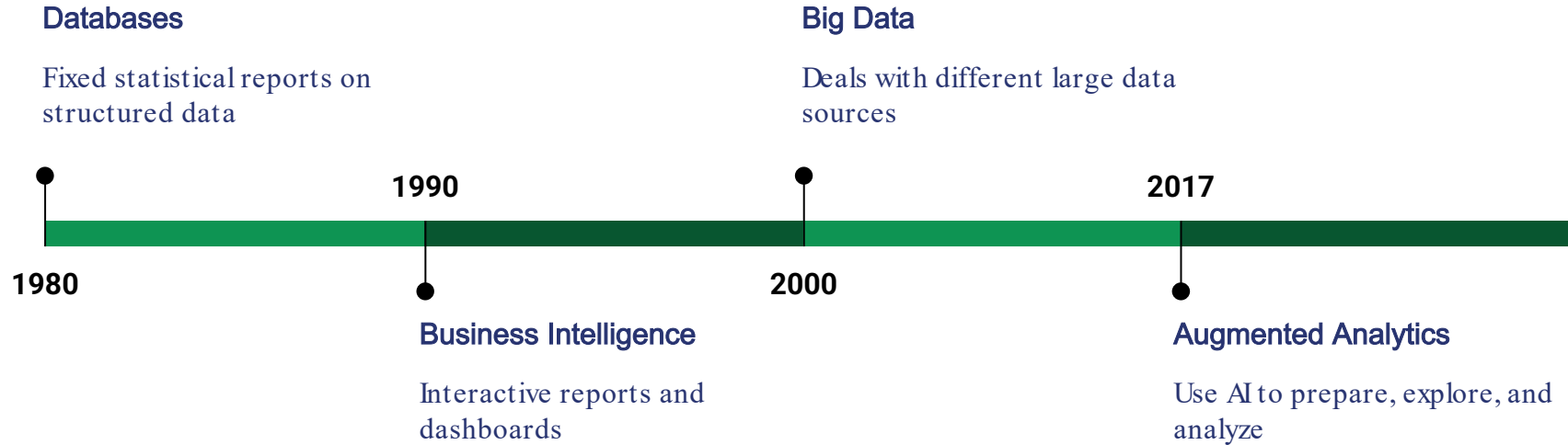
## Importance and Benefits

- Improve: *If you can't measure it you can't improve it*
- Cost savings: *DeepMind saved Google 40% of cooling costs*
- Inform decisions: *Football clubs use players' data during recruitment*
- Innovation: *Startups leverage market data to exploit their niche demands*





# History of Data Analytics



# Types of Data Analytics



## **Descriptive Analytics** (*Past- What*)

Summarizes historical data to provide insights into past events or current situations



## **Diagnostic Analytics** (*Reason- Why*)

Examines data to identify the root causes or reasons behind specific events or outcomes



## **Predictive Analytics** (*Future - What will*)

Uses historical data and statistical models to forecast future trends or outcomes



## **Prescriptive Analytics** (*Advice- What to do*)

Combines data analysis with business rules and mathematical models to recommend optimal courses of action (can be used to automate decision making)



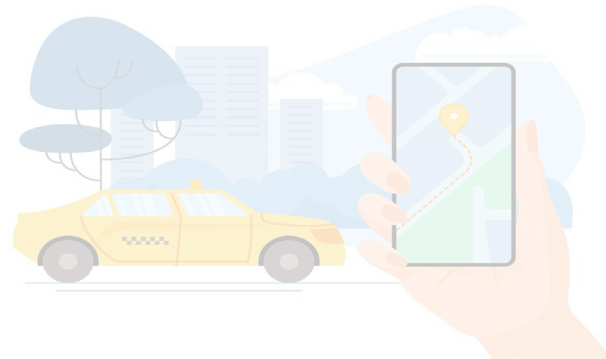
# RideSharing Driver Analytics

## Descriptive Analytics (*Past- What*)

- Analyzing historical data to know peak hours when demand for rides is highest
- Visualizing the geographic distribution of ride requests across different locations
- Summarizing the average wait times and ride durations for different time periods

## Diagnostic Analytics (*Reason- Why*)

- Investigating the reasons behind longer wait times during certain hours or locations
- Analyzing the impact of weather, traffic patterns, or special events on ride demand
- Finding the causes of rider cancellations or low driver ratings during specific times



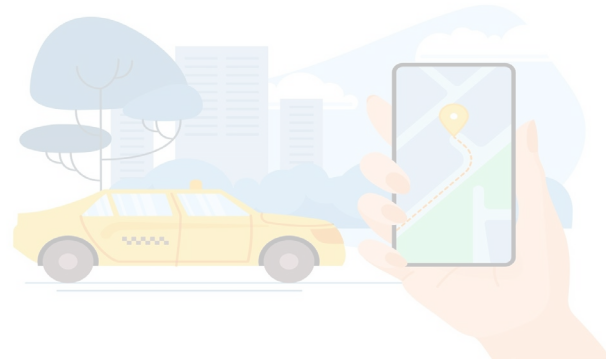
# RideSharing Driver Analytics

## Predictive Analytics *(Future - What will)*

- Developing models using historical data to forecast demand based on time of day, day of the week, and other relevant factors
- Predicting the likelihood of high demand during upcoming events or holidays
- Estimating driver earnings based on predicted ride volumes and pricing

## Prescriptive Analytics *(Advice- What to do)*

- Recommending the best locations for drivers based on predicted demand
- Providing guidance on the optimal routes or travel paths to minimize idle time
- Optimizing driver shift assignments to maximize earnings and meet demand



# Data Analytics Examples



## Latency for Most Common Video conferencing Platforms in the Kingdom

<https://www.meqyas.sa/>

Source: Based on analysis by SamKnows Ltd. for the period from January to March 2023

1

First Place



92 ms



220 ms



88 ms



31 ms

2

Second Place



95 ms



222 ms



90 ms



39 ms

3

Third Place



102 ms



226 ms



93 ms



67 ms

4

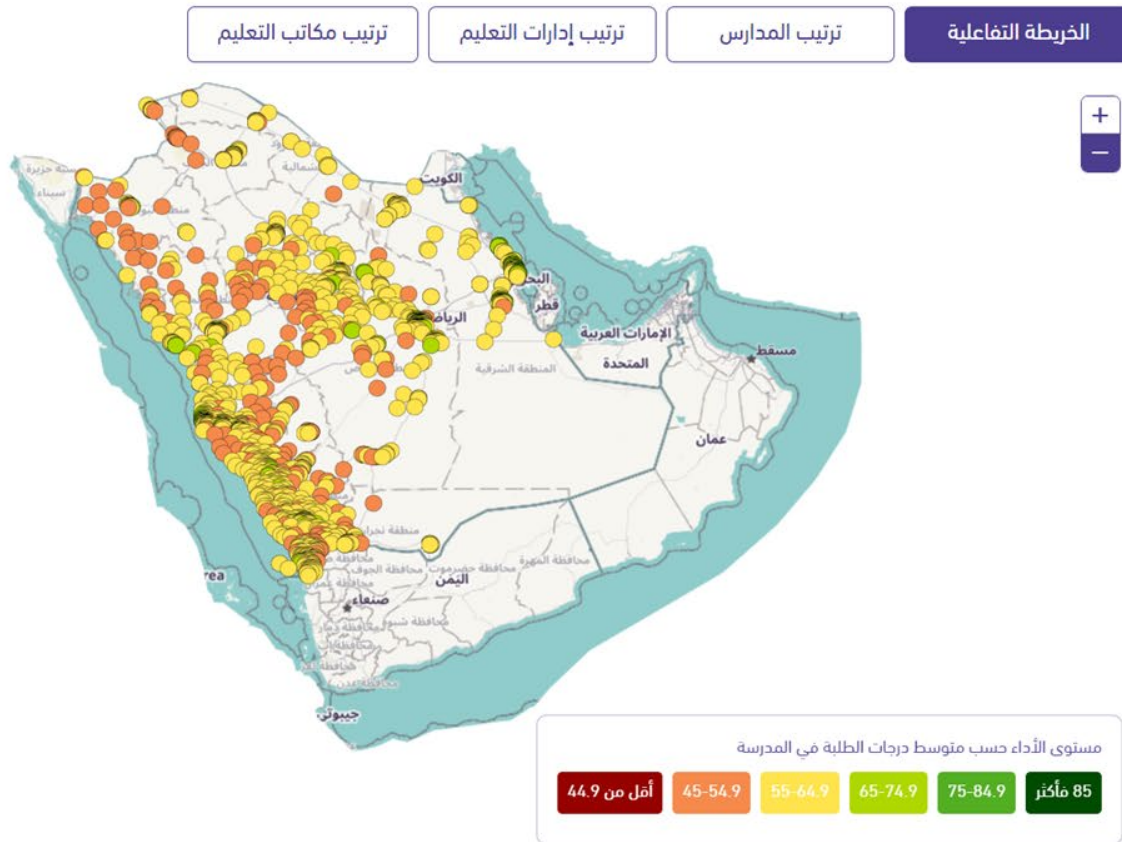
Fourth Place



110 ms



82 ms



## خيارات البحث

العام الدراسي

1442

1443

1444

نوع الاختبار

قدرات علمي

تحصيلي نظري

تحصيلي عملي

قدرات نظري

مستوي الأداء

65-74.9

75-84.9

85 فأكثر

الكل

أقل من 44.9

45-54.9

55-64.9

الجنس

بنات

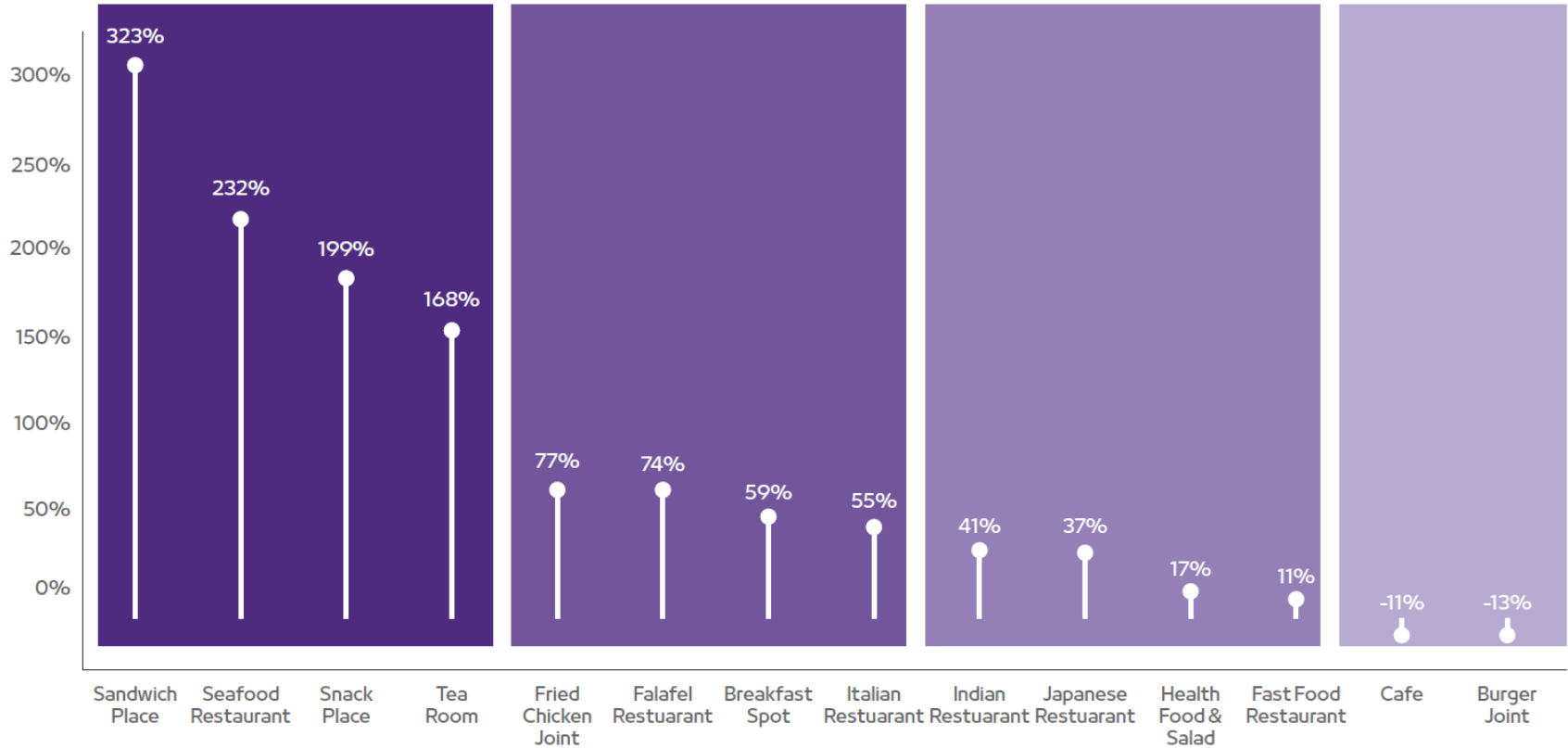
بنين

إعادة تعيين

تطبيق

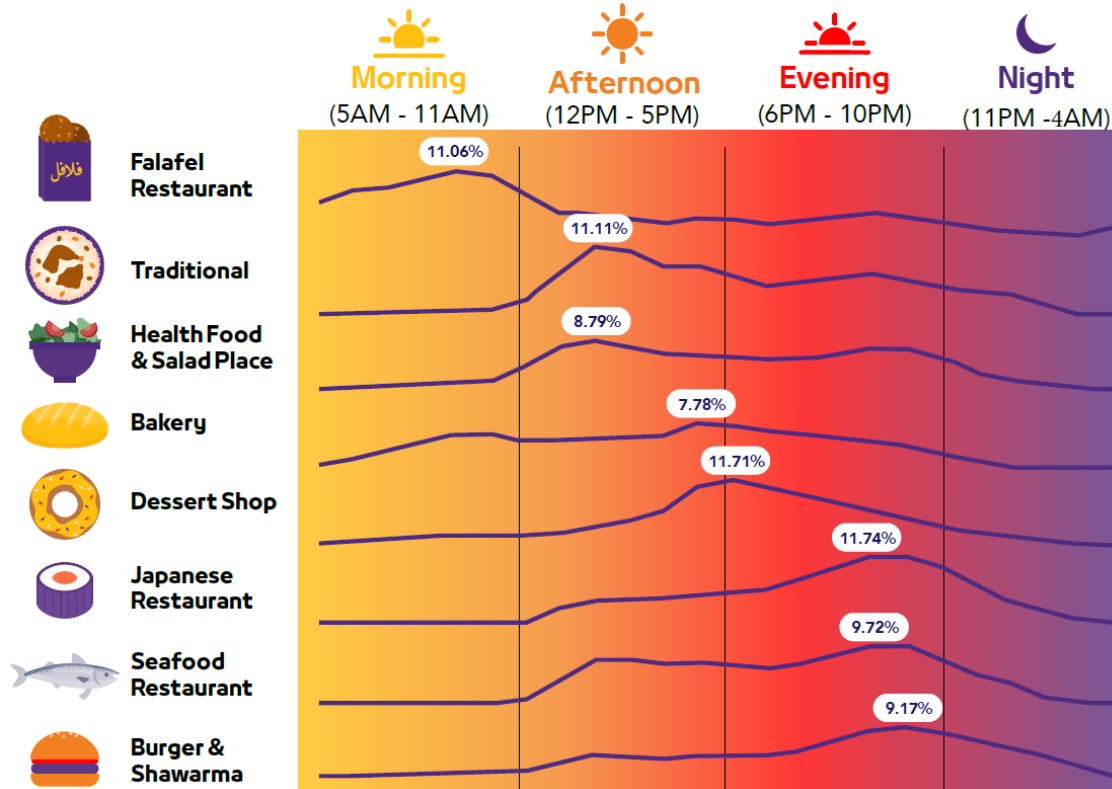
# Food & Beverages Delivery Report

Growth in number of New Branches between 2022 and 2023

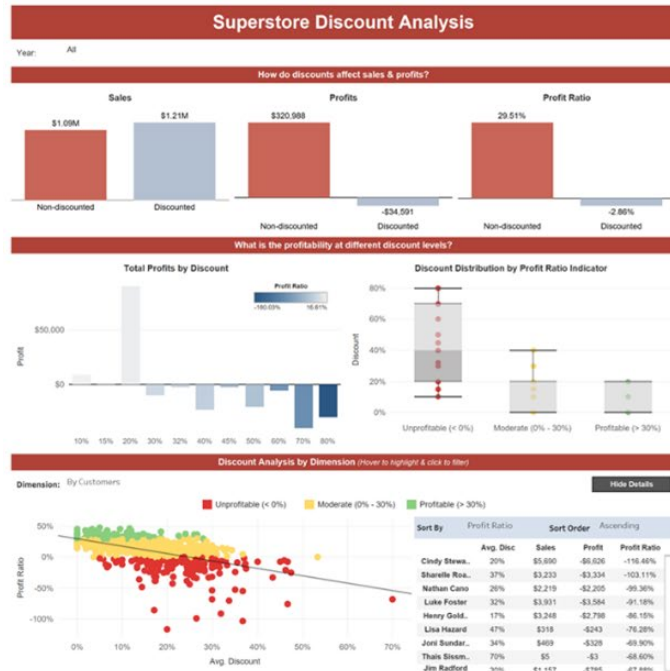




# Food & Beverages Delivery Report



# Data Analytics Tools



# Data Analytics Use Cases



# Data Analytics Use Cases



## Education

- Admission Screening
- Performance Monitoring
- Aptitude Testing



## Finance

- Fraud Detection
- Risk Analysis
- Illegal Activity Detection



## Manufacturing

- Predictive Maintenance
- Customer Satisfaction
- Improving Products and Services



## Agriculture

- Weather Prediction
- Natural Disaster Prediction
- Increase Yield Quality and Quantity



# Data Analytics Use Cases



## Government

- Market Evaluation
- Improve Economy
- Predict Cyber Attacks



## Transportation

- Improve Traffic
- Improve Delivery
- Logistic Services



## Security

- Faster Response
- Crime Prediction
- Less Cybercrimes



## Energy

- Demand Prediction
- Resource Utilization
- Tampering Detection



## Healthcare

- Practitioner Evaluation
- Anomaly Detection
- Pandemic Prediction



# Data Analytics Workflow



# Data Analytics Workflow



The following steps apply to all types of data analytics:

## Ask

Define the business problem or question that needs to be answered

## Prepare

Collect, clean, and organize the relevant data from various sources into a format suitable for processing

## Process

Apply appropriate data processing techniques, such as data transformation, feature engineering, and data modeling

## Analyze

Use statistical techniques, machine learning algorithms, and visualization tools to uncover patterns, trends, and insights from the data

## Act

Communicate the findings, insights, and recommendations to stakeholders and implement data-driven decisions to drive business value



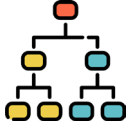
# Data Analytics Considerations

- Data Accessibility
- Data Bias
- Data Quality
- Data Quantity
- Data Diversity
- Data Privacy
- Data Security
- Selecting the Right Analytical Tool





# Data Structures



## Structured

- Tabular Data
- Relational Databases



## Semi-Structured

- XML
- JSON
- NoSQL



## Unstructured

- Text
- Audio
- Images
- Videos
- Webpages



# Data Types

- Qualitative(Categorical)
  - Nominal: Gender, Marital Status, Color, Brand, Favorite Sport
  - Ordinal: Education Level, Customer Rating, Income Level
- Quantitative(Numerical)
  - Discrete: Number of Children, Number of floors, Number employees
  - Continuous: Height, Weight, Price, Temperature, Distance



# Data Analytics Case Study



# Case Study: Aramco's Data Analytics Journey

- **Aramco**, a leading energy company, recognized the potential of data analytics to drive efficiency and innovation
- They embarked on **data-driven transformation** , leveraging the power of big data and advanced analytics
- Aramco uses big data analytics **to improve efficiency, reduce emissions and enhance operations**
- By leveraging data analytics and following the workflow, Aramco achieved significant improvements in areas such as **asset reliability, energy efficiency, and operational excellence.**



# Case Study: Aramco's Data Analytics Journey

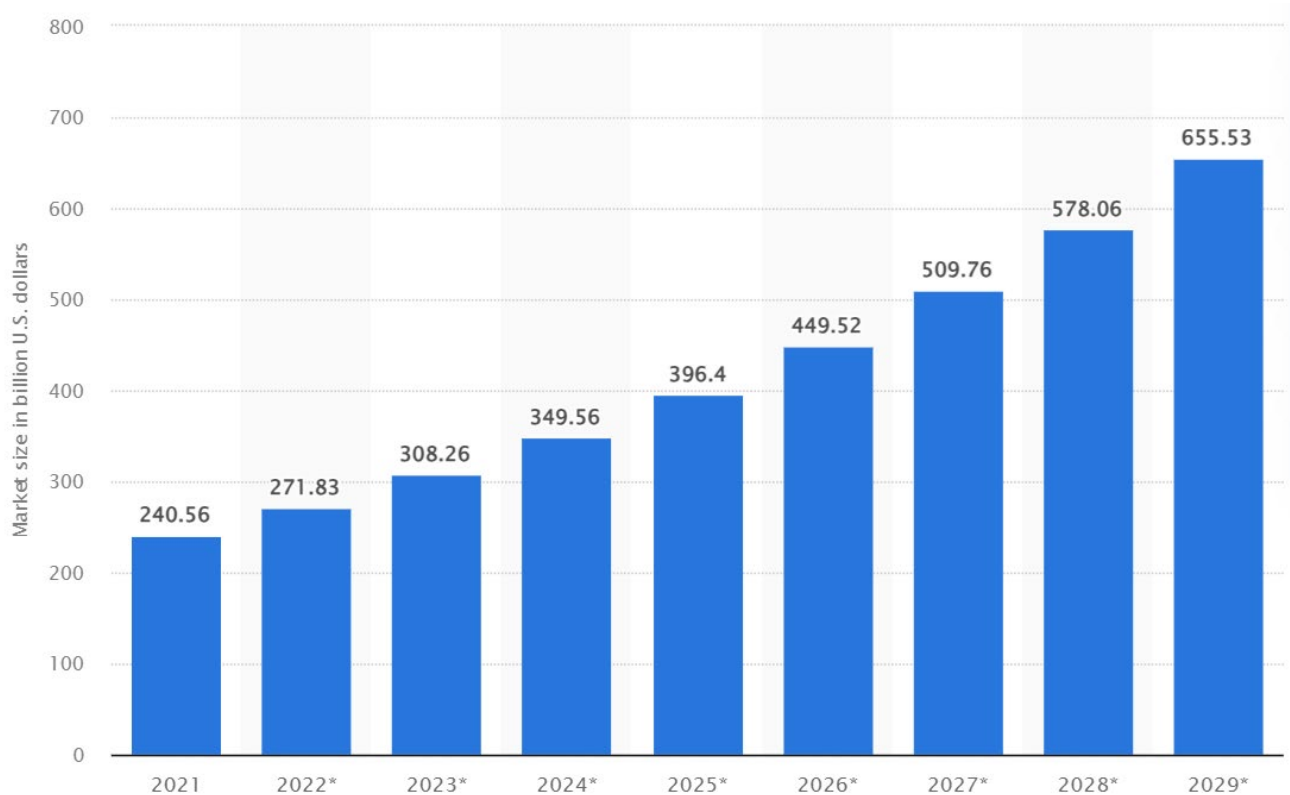
- **Descriptive Analytics:** Aramco collects data from various sources including sensors, engineering drawings, and inspection data to gain insights into historical data, such as production rates, current trends, and equipment performance.
- **Diagnostic Analytics** By analyzing sensor data and operational parameters, Aramco diagnosed the root causes of issues and inefficiencies.
- **Predictive Analytics:** Aramco uses data analytics to predict equipment failure, which helps them to take preventive maintenance actions and avoid downtime.
- **Prescriptive Analytics:** Aramco uses data analytics to optimize production and reduce flaring. This helps them to make data-driven decisions that improve their bottom line and environmental impact.



# What is Next?



# Forecasted Data Analytics Market Size



# Data Job Roles

- **Data Engineer** builds and maintains the data infrastructure, including data pipelines, data warehouses, and data lakes, to ensure efficient and reliable data flow within an organization
- **Data Architect**: designs and oversees the overall data architecture, including data models, data governance policies, and data standards, to ensure that data is organized, integrated, and accessible across the enterprise
- **Data Analyst** collects, cleans, and analyze data to derive insights and present them in a meaningful way to support data-driven decisionmaking
- **Data Scientist** applies advanced statistical and machine learning techniques to extract insights and patterns from complex data sets, build predictive models, and develop data-driven solutions to business problems





# Extra Resources and References

- [Data Analytics Course from Google](#)
- [SDAIA Data Analytics for Executives Report](#)
- [Saudi Open Data Platform](#)
- [Open Data Usage Examples and Success Stories](#)
- [Data Analytics in the Healthcare Sector](#)
- [Importance of Data Analytics FinancesOnline Post](#)
- [Aramco's Data Analytics Journey](#)
- [DeepMind's Cost Saving using Data Analytics](#)
- [How Data is Transforming Football Recruitment](#)
- [How Uber Leverages Data Analytics](#)
- [Data Storytelling using Power BI](#)



# Thank You



**SDAIA**

الهيئة السعودية للبيانات  
والذكاء الاصطناعي  
Saudi Data & AI Authority