SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
Accredited by NAAC-UGC with 'A++' Grade (Cycle III) &
Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)
COIMBATORE-641 035, TAMIL NADU

DEPARTMENT OF COMPUTER APPLICATIONS

19CAE716 – DATA SCIENCE

UNIT – I: INTRODUCTION TO DATA SCIENCE

TOPIC: DATA SCIENCE – INTRODUCTION & DEFINITION

What is Data?

- ✓ Data **is a collection of information** gathered by **observations, measurements, research or analysis**.
- ✓ They may consist of facts, numbers, names, figures or even description of things.
- ✓ Data is organized in the form of graphs, charts or tables.

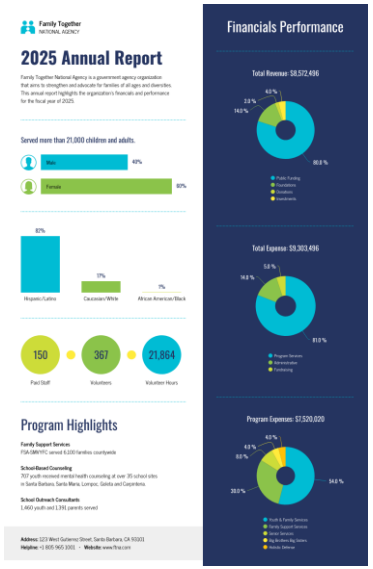
	A	B	C	D	E
1	Last Name	Sales	Country	Quarter	
2	Smith	$16,753.00	UK	Qtr 3	
3	Johnson	$14,808.00	USA	Qtr 4	
4	Williams	$10,644.00	UK	Qtr 2	
5	Jones	$1,390.00	USA	Qtr 3	
6	Brown	$4,865.00	USA	Qtr 4	
7	Williams	$12,438.00	UK	Qtr 1	
8	Johnson	$9,339.00	UK	Qtr 2	
9	Smith	$18,919.00	USA	Qtr 3	
10	Jones	$9,213.00	USA	Qtr 4	
11	Jones	$7,433.00	UK	Qtr 1	
12	Brown	$3,255.00	USA	Qtr 2	
13	Williams	$14,867.00	USA	Qtr 3	
14	Williams	$19,302.00	UK	Qtr 4	
15	Smith	$9,698.00	USA	Qtr 1	
16					

What is Data Science?

- ✓ Data science is the study of data to extract meaningful insights for business.
- ✓ It is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data.
- ✓ This analysis helps data scientists to ask and answer questions like what happened, why it happened, what will happen, and what can be done with the results.

1

2

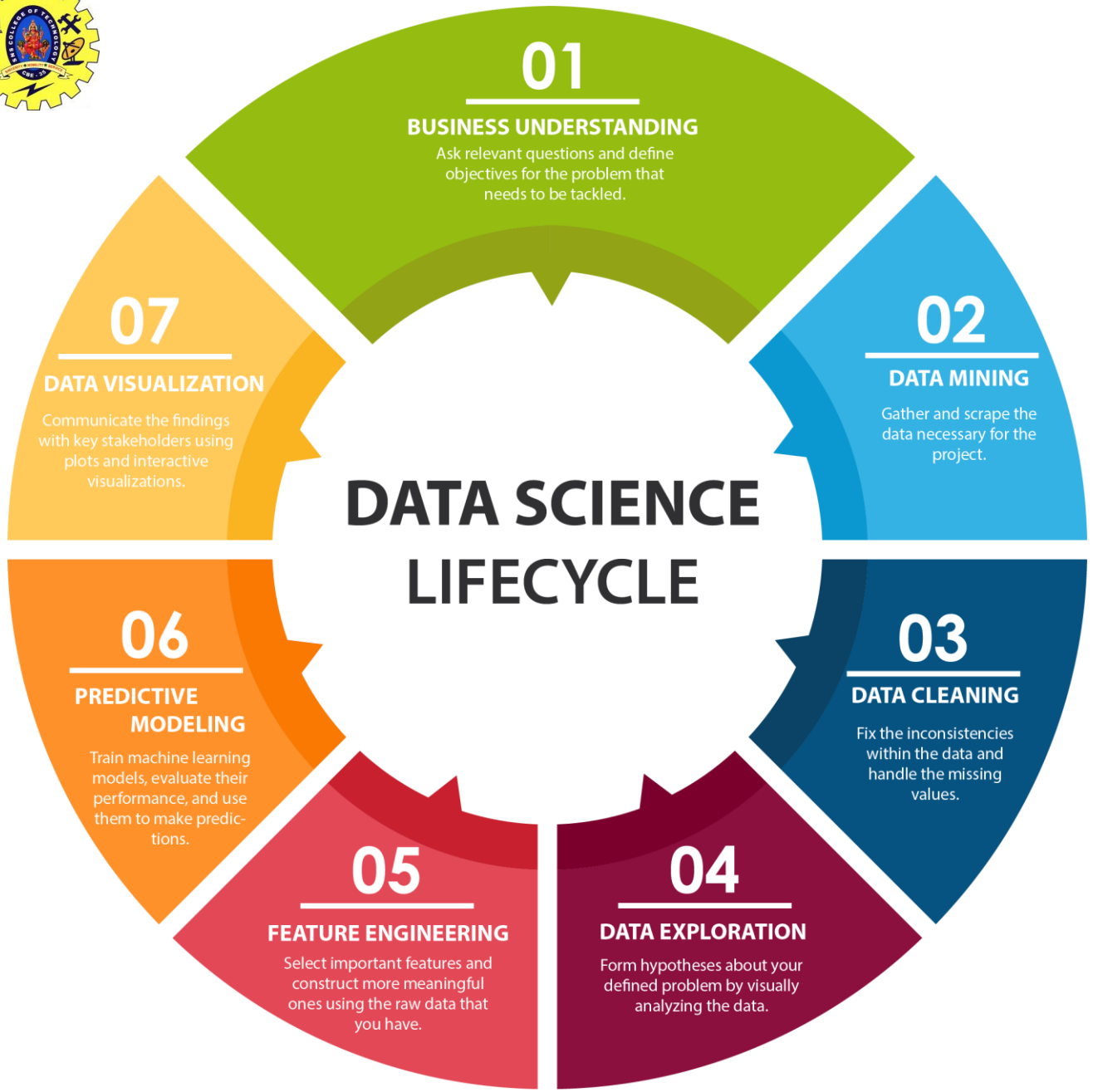


5

3

EXPLORE DATA

4



Data Science Life Cycle

01 Business Understanding

02 Data Mining

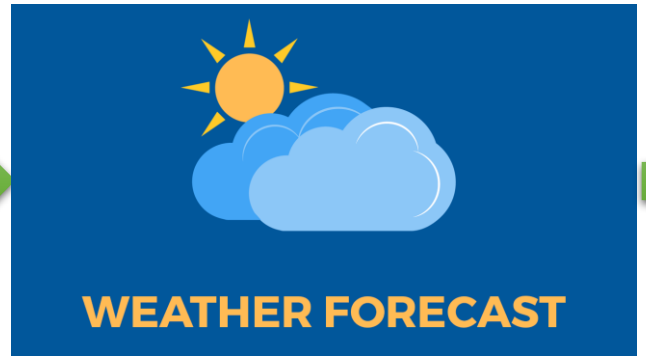
03 Data Cleaning

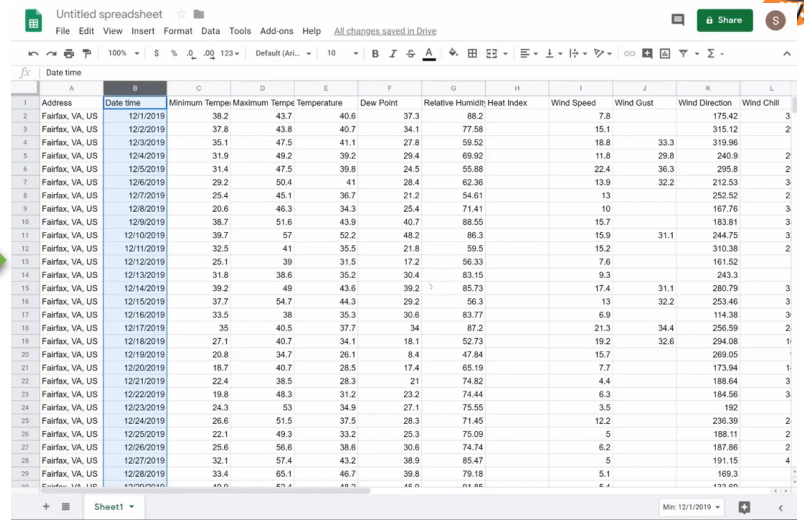
04 Data Exploration

05 Feature Engineering

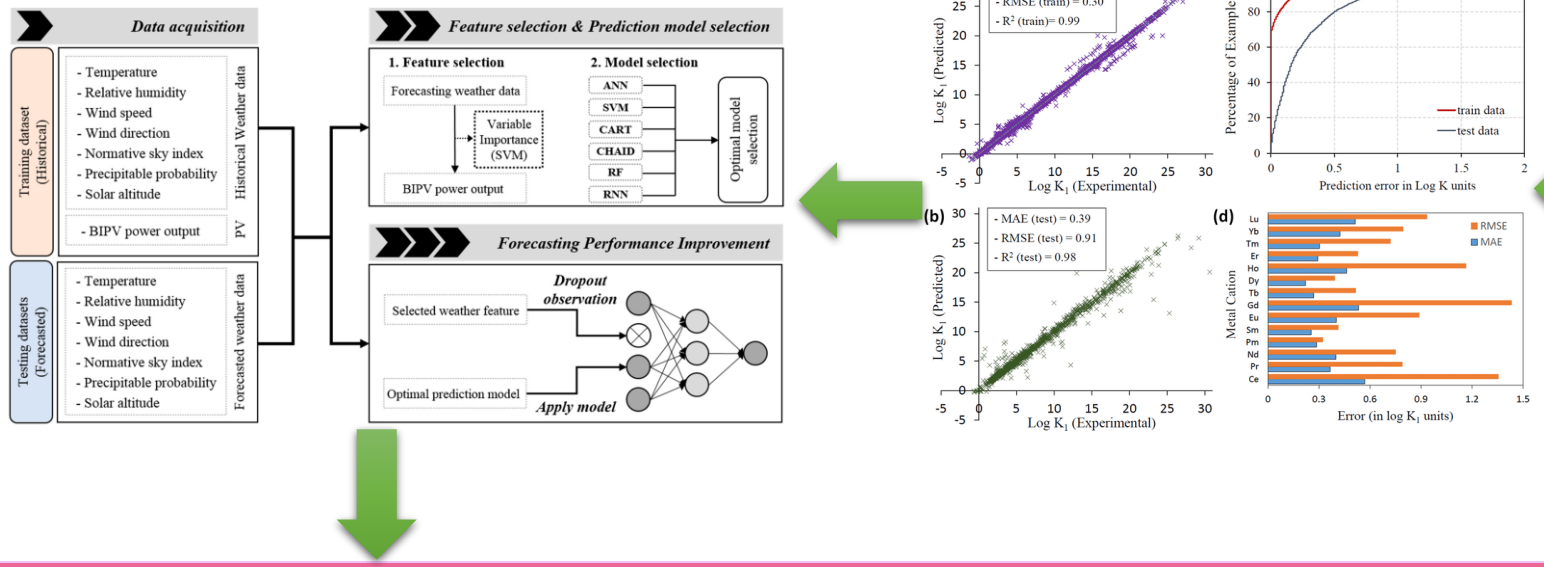
06 Predictive Modeling

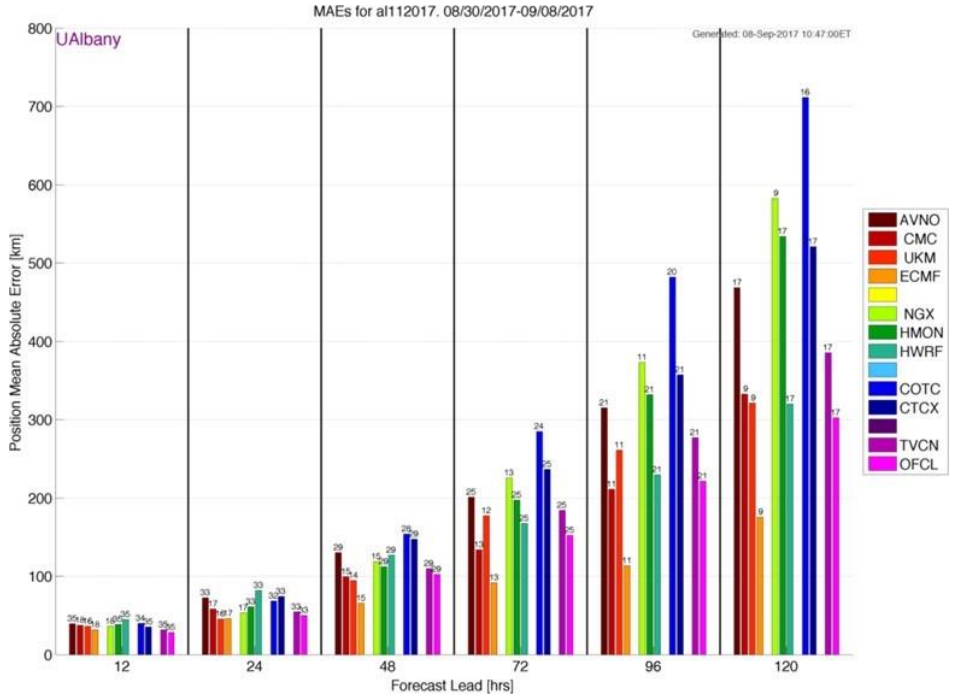
07 Data Visualization

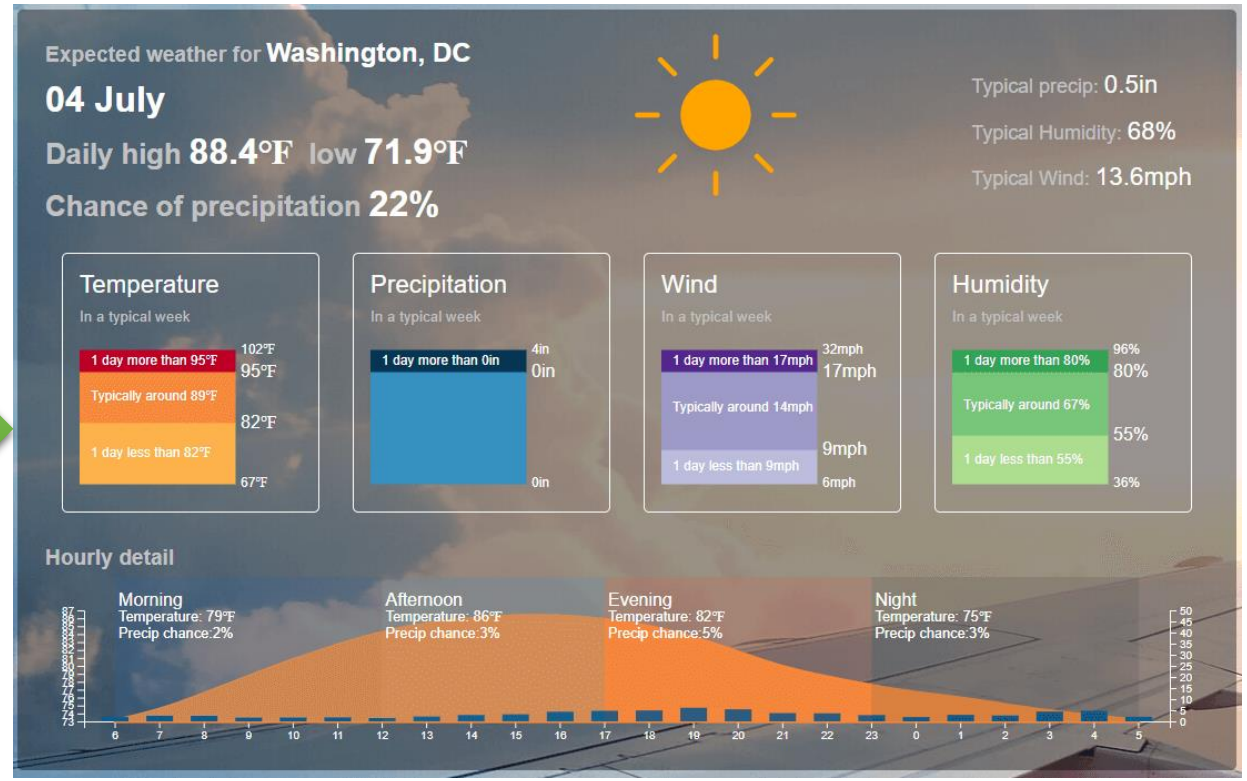




	Unnamed: 0	date	day	desc	humidity	precip	temp	wind
0	0	20 MAY	Tonight	Partly Cloudy	89%	10%	--22°	WSW 14 km/h
1	1	21 MAY	Thu	Partly Cloudy	59%	20%	34°22°	W 15 km/h
2	2	22 MAY	Fri	Partly Cloudy	56%	10%	35°23°	NW 16 km/h
3	3	23 MAY	Sat	Partly Cloudy	55%	20%	36°23°	WNW 15 km/h
4	4	24 MAY	Sun	Partly Cloudy	56%	20%	35°23°	WNW 15 km/h
5	5	25 MAY	Mon	PM T-Storms	61%	60%	35°23°	WNW 14 km/h
6	6	26 MAY	Tue	PM T-Storms	64%	80%	34°23°	WSW 12 km/h
7	7	27 MAY	Wed	PM T-Storms	69%	80%	32°22°	WSW 12 km/h
8	8	28 MAY	Thu	PM T-Storms	73%	80%	31°22°	W 14 km/h
9	9	29 MAY	Fri	PM T-Storms	73%	80%	31°22°	WSW 12 km/h
10	10	30 MAY	Sat	PM T-Storms	74%	80%	30°22°	W 12 km/h
11	11	31 MAY	Sun	Scattered T-Storms	72%	60%	31°22°	W 14 km/h
12	12	1 JUN	Mon	T-Storms	72%	80%	31°22°	WSW 13 km/h
13	13	2 JUN	Tue	T-Storms	73%	80%	29°22°	WSW 15 km/h
14	14	3 JUN	Wed	Scattered T-Storms	74%	60%	29°21°	WSW 15 km/h

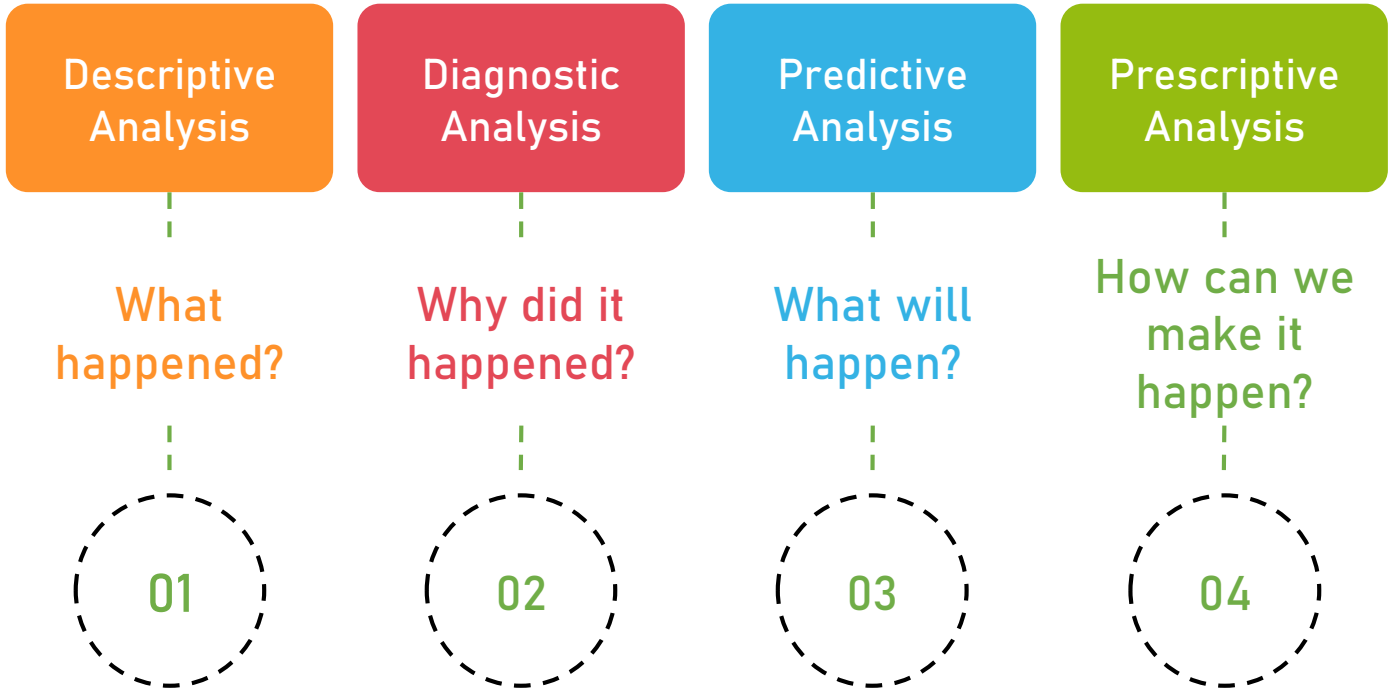



MAEs for al112017. 08/30/2017-09/08/2017
UAlbany
Generated: 08-Sep-2017 10:47:00ET
Position Mean Absolute Error [km]
Forecast Lead [hrs]
AVNO
CMC
UKM
ECMF
NGX
HMON
HWRF
COTC
CTCX
TVCN
OFCL


Expected weather for Washington, DC
04 July
Daily high 88.4°F low 71.9°F
Chance of precipitation 22%
Typical precip: 0.5in
Typical Humidity: 68%
Typical Wind: 13.6mph
Temperature
In a typical week
1 day more than 95°F
Typically around 89°F
1 day less than 82°F
Precipitation
In a typical week
1 day more than 0in
Wind
In a typical week
1 day more than 17mph
Typically around 14mph
1 day less than 9mph
Humidity
In a typical week
1 day more than 80%
Typically around 67%
1 day less than 55%
Hourly detail
Morning
Temperature: 79°F
Precip chance:2%
Afternoon
Temperature: 86°F
Precip chance:3%
Evening
Temperature: 82°F
Precip chance:5%
Night
Temperature: 75°F
Precip chance:3%

Types of Analysis

- ✓ Analysis of data is a vital part of running a successful business.
- ✓ When data is used effectively, it leads to better understanding of a business's previous performance and better decision-making for its future activities.
- ✓ There are many ways that data can be utilized, at all levels of a company's operations.



Descriptive Analysis

What happened?

01

Purpose: Descriptive analysis involves summarizing and presenting key features of a dataset.

Methods: This includes measures of central tendency (mean, median, mode), measures of variability (range, standard deviation), and graphical representations (histograms, pie charts, bar charts).

Example: Analyzing the average order value and the distribution of products sold on Flipkart in a given month.

Diagnostic Analysis

Why did it happened?

02

Purpose: Diagnostic analysis focuses on understanding the causes of observed phenomena.

Methods: Root cause analysis and correlation analysis are common. It involves identifying patterns and relationships to explain why certain events occurred.

Example: Investigating a sudden drop in customer satisfaction scores on Flipkart and identifying the factors contributing to it.

Predictive Analysis

What will happen?

03

Purpose: Predictive analysis involves forecasting future trends and outcomes based on historical data.

Methods: Statistical modeling, machine learning algorithms, and time series analysis are used to make predictions.

Example: Using customer purchase history to predict future buying behavior and tailor marketing strategies accordingly.

Prescriptive Analysis

How can we make it happen?

04

Purpose: Prescriptive analysis goes beyond predicting outcomes by providing recommendations for actions to optimize results.

Methods: Optimization algorithms, simulation models, and decision analysis are used to suggest the best course of action.

Example: Recommending personalized product recommendations on Flipkart based on a combination of customer preferences, behavior, and trends.