Introduction to Data Analytics:

Why do we need data analytics?

- is today's business world, having an ability to utilize data and analysis place a vital role for decision making and business-related things.

- Data and information are unpredictable, but data analytics helps to analyze the data we have using trends, patterns and information. for better profit.

- Managing cash flow

- what you cannot attract, you can't control

Data Analytics - the process of examining and analyzing raw data sets to:

1. Draw conclusion

2. Derive information

3. Improve business, products and services.

Data Analytics Process Flow:

1. Define goals

2. Identify measurable metrics

3. List, collect and extract data from sources

4. Explore and analyze data

5. Interpret and visualize data

6. Infer data for decision-making

Data Analytics Life Cycle

1. Discovery - learn about business domain and assess available resources

2. Data Preparation - Execute ELT (extract, load, transform)

3. Model Planning - Identify techniques and data to understand variables relationship

4. Model Building - develop datasets for testing, training and production

5. Communicate Results - Identify key findings, business values, and develop narrative to convey to stakeholders

6. Operationalize - deliver final reports, briefs, codes, and technical documents.

Types of Data Analytics

1. Descriptive Analytics - What happened?

- Helps you get historical data

- provides summarized information of facts.

Data Aggregation and Data Mining

- Data Aggregation is the process of gathering and summarizing information, tools used are ms excel, matlab, spss, and stata

- Examples are company report.

2. Diagnostic Analytics - Why did this happen?

- Deeper look at the data to understand the root causes

- Provides understanding about relationships and consequences

- Drill Down, Data Discovery, Data Mining, and Correlation

Data Mining - the process of sorting large data sets, helps identify patterns and establish relationships to solve problems.

Most concerned with day to day operation

3. Predictive Analytics - What will happen?

- Use to predict future outcomes in terms of probability an event to occur.

- Sentiment analytics using model

- identifies the customers to be included in a promotional campaign for maximizing response.

- Weather forecast, and such.

- Utilizing machine learning analysis

- Such as Random forest, svm

- python, r, and rapid minor are the famous environment

4. Prescriptive Analytics - How can we make it happen?

- Not everyone use this.

Examples of type of analytics:

1. Diagnostic Analytics, Netflix, due to squid game, the revenue of netflix spiked very high.

2. Predictive Analytics - Create a season 2 for it. since it's what caused the all time high.

Data Analytics Benefits: Decision-Making

1. Data analytics helps you define your target audience

- true to, like using data of what really is trending can make the company bloom

- Shopping patterns, Demographic, Time of purchase

Examples

1. Footwear and clothing worth around 642 billion are retruned to stores every year,

this happens as consumers miss important information during the purchase,

so what can we do?

find out which information are mostly being forgotten, and then the company should implement in applying it so that the return chances will be minimized, by posting product specifications, or posting product videos, and such

Case Study: Amazong

1. Amazong uses predictive analysis for chossing the warehouse closest to the consumers reducing the shipping cost by 10 to 40 percent.

- cutting the delivery time to 1 day which is a very huge thing since time is gold.

2. Amazon also use it for prices, based on website activity, competitor

DATA VISUALIZATION TOOLS

1. Power BI

2. Tableau

3. Logi

The goals of data analytics based on my understanding:

1. identify outliers and trends

2. reduce cost for company

3. explore ways to increase profit

4. make the company bloom as versatile as possible.

DATA ANALYTICS TOOLS:

- used when data are unstructured

Dealing with Different Types of Data

Terminologies in Data Analytics:

1. Observation - a single row of data from database

- is the unit of analysis on which the measurements are taken.

Database table has; variables and rows

2. Data Sampling - statistical analysis technique used to select, manipulate, and analyze a representative subset of data points.

Goal: identify patterns and trends in a larger dataset

3. Dataset - A collection of data or the total data captured about a particular use case

- medical records, insurance records, loan application approval records

4. Prediction - move from what has happened the asses what will happen in the future

TYPES OF DATA

1. Structured - data processed, stored, and retrieved in fixed format

example is employee file

2. Semi-Strcutred - data that contsins both structured and unstructured

example is csv file

3. Unstructured - data lacks specific form of structure

example is an email

ANALYZING UNSTRUCTURED DATA

QUALITATIVE AND QUANTITATIVE DATA

1. Qualitative data - classifications of objects i based on attributes and properties

- unstructured data.

- nonstatistical

- decriptive in nature

- asks why

Subgroups are:

1. Nominal Data - Unordered data which an order is assigned, in relation to other named categories

example are: pass or fail

2. Ordinal Data - ordered data which is assigned to categories in a ranked fashion.

example are ratings, good excelelent and all.

2. Quantitative data - data is measured numerically.

- structured data

- statistics and numbers

Subgroups are:

1. Discrete Data - Only takes certain values

example: employment status

2. Continuous Data - Takes any value within a specific range which can be infinite.

example: the height of a person, shared revenue of company

DATA LEVELS OF MEASUREMENT

- a classification that describes the nature of information within the values assigned to variables.

4 data labels:

1. Nominal - the numbers in the variable are used to classify the data

- such as words, letter, and alpha

- example are F for female, M for Male

2. Ordinal - depicts ordered relationship among the variables observations.

examples, 100 for gold, 92 for silver, 81 for bronze

3. Interval - classifies and orders the measurements, also specifies that he distances between each interval on the scale are equivalent.

example temperature in centigrade

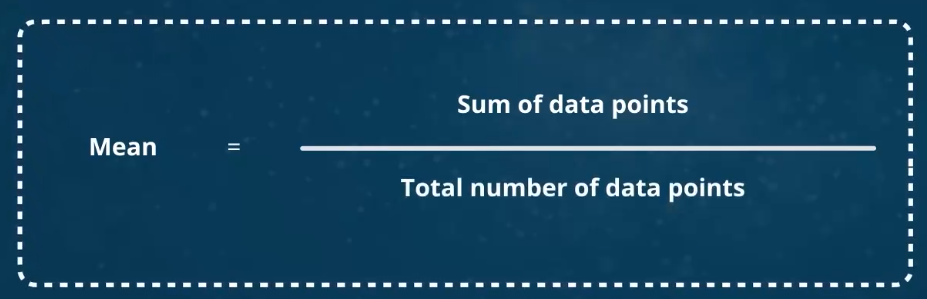
4. Ratio - Obsercations can have a value of zero, has a similar properties as the interval level of measurements

NORMAL DISTRIBUTION OF DATA

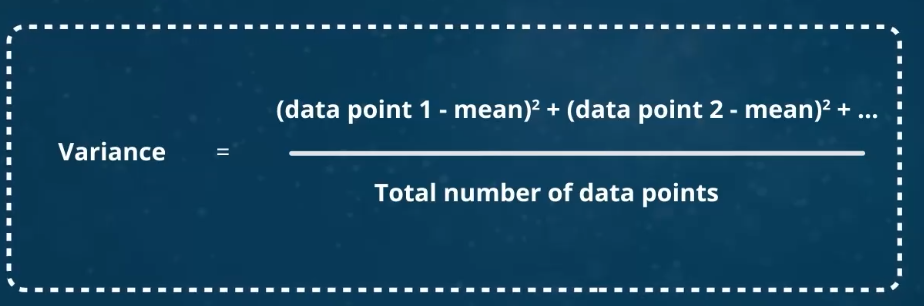
1.Gaussian distribution or bell curve

Statistical Parameters:

1. mean - average of all the data points for a given set of data, add all the data points then divide to number of points



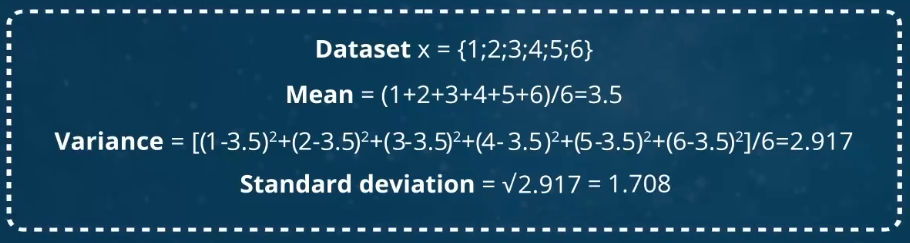
2. variance - looks at all the data points and then determines their distribution



3. standard deviation - tells you how tightly all the variance example are clustered around the mean



Example:



**DATA VISUALIZATION FOR DECISION MAKING**

Data Visualization:

this is criticial to pique the attention of everyone.

Commonly used visualization:

1. Heatmap - uses a warm to cool color spectrum, show strength of relationships with color.

2. Frequency Dsitribution Plot - is the mumber of times the event occurs in an observation. an umbrella term

- Table

- Histogram

- Line Graph

- Dot Plot

- Pie Chart

3. Swarm Plot - plots in a single dimension to show frequency and such, enables to separate all overlapping points making each point visible

Importance of Data Visualization in Analytics

1. Provides access to know data trends, outliers, and patterns. Hence help business in the long run.

2. Where to cut costs or how to improve operational processes.

3. Unorganize is unhelpful

Exploratory Data Analytics

1. first step towards modelling, get detailed insights into the dataset, detect if any outliers is present

Data Visualization Tools: Part 1

1. Tableau - it is simple to use and produces better looking interactive visualization

2. FusionCharts -

3. Highcharts - Anyone can view this, might be good for beginner

4. DataWrapper -

5. Plotly -

6. SiSense -

7. Power BI - Esp for microsoft

8. Looker - has templates

9. Domo - focus in social collaboration

10. Qlik - clean interface, highly customizable setup

11. board - full featured business intelligence system

Languages in Data Visualization:

1. Scala

2. Python - matplotlib, seaborn

3. R - base grapphics, lattice graphics, grid graphics

4. javascript -

5. java - java2d, java3d, java advance

Dashboard-Based Visualization

1. all the important parameters has in different visualizations

2. should be interactive

3. should be customizable

steps to make this:

1. analyze your target audience

2. identift the business parameters such as KRA, KPI, KPA, SLA

3. Know the end goal of the dashboard

4. Get hands-on in developing the dashboard

5. imporvised based on current situation

BI and Visualization Trends

1.