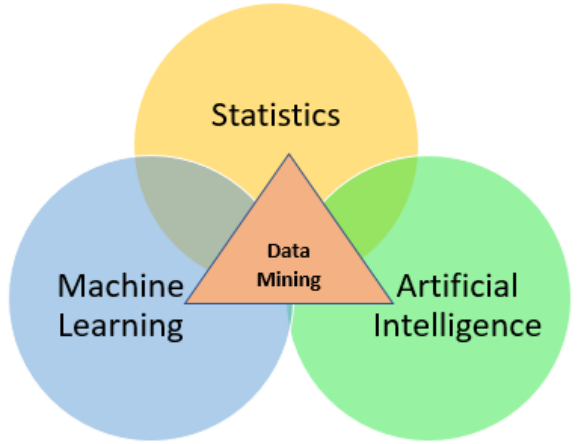


Data Analytics

Tilani Gunawardena
PhD(UNIBAS), BSc.Eng(Pera), FHEA(UK), AMIE(SL)
2019/11/22

Why?


facebook
Social Analytics
amazon
Improving E-commerce experience
UBER
Optimizing Rides
Bank of America
Increasing Customer experience
Airbnb
Improving Searches
Spotify
Music Recommendation

Data

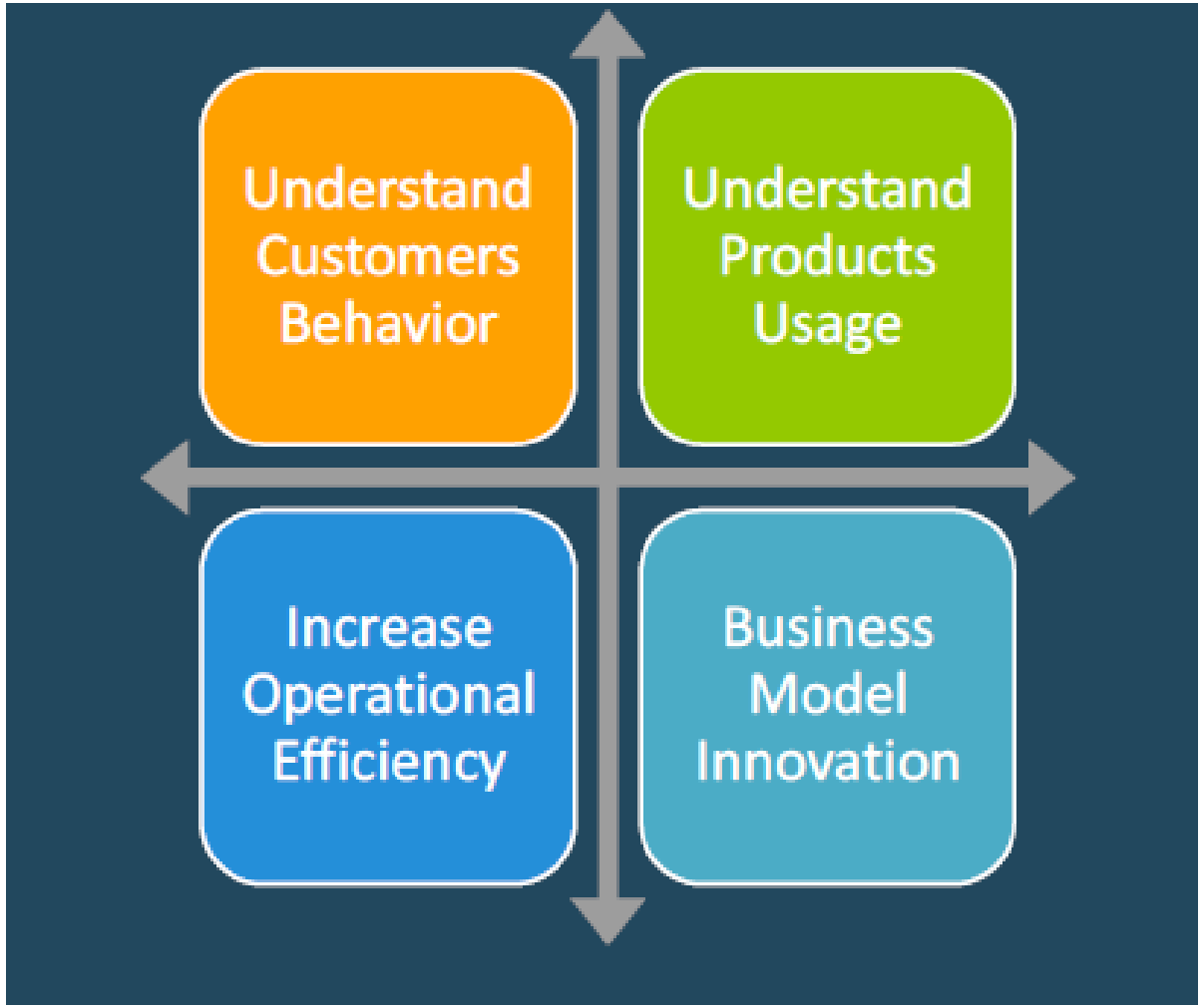
- Data- Data is a set of values of qualitative or quantitative variables. It is information in raw or unorganized form. It may be a fact, figure, characters, symbols etc.
- Information- Meaningful or organized data is information, comes from analyzing data

Data Analytics

- Analytics- Analytics is the discovery , interpretation, and communication of meaningful patterns or summery in data.
- Data Analytics (DA) is the process of examining data sets in order to draw conclusion about the information it contains.
- Analytics is not a tool or technology, rather it is the way of thinking and acting on data.

Data on its own is useless unless you can make sense of it!

Primary Focus Areas for Analytics



Data Analytics - Example

- Business analytics
- Risk
- Fraud
- Health
- Web

Business Analytics

BUSINESS NEED

- The Business environment today is more complex than ever before.
- Businesses are expected to be diligently responsive to the increasing demands of customers, various stakeholders and even regulators.

SOLUTION

- Organizations have been turning to the use of analytics.
- More than 83% of Global CIOs surveyed by IBM in 2010 singled out Business Intelligence and Analytics as one of their visionary plans for enhancing competitiveness.

GOAL

In most cases the primary objective of an organization that seeks to turn to analytics is:

- Revenue/Profit growth
- Optimize expenditure

The Process of Statistical Analysis

When we have resource constraints, Statistical Analysis enables us to make quantitative inferences based on an amount of information we can analyze (a sample).

Data Analytics Life Cycle

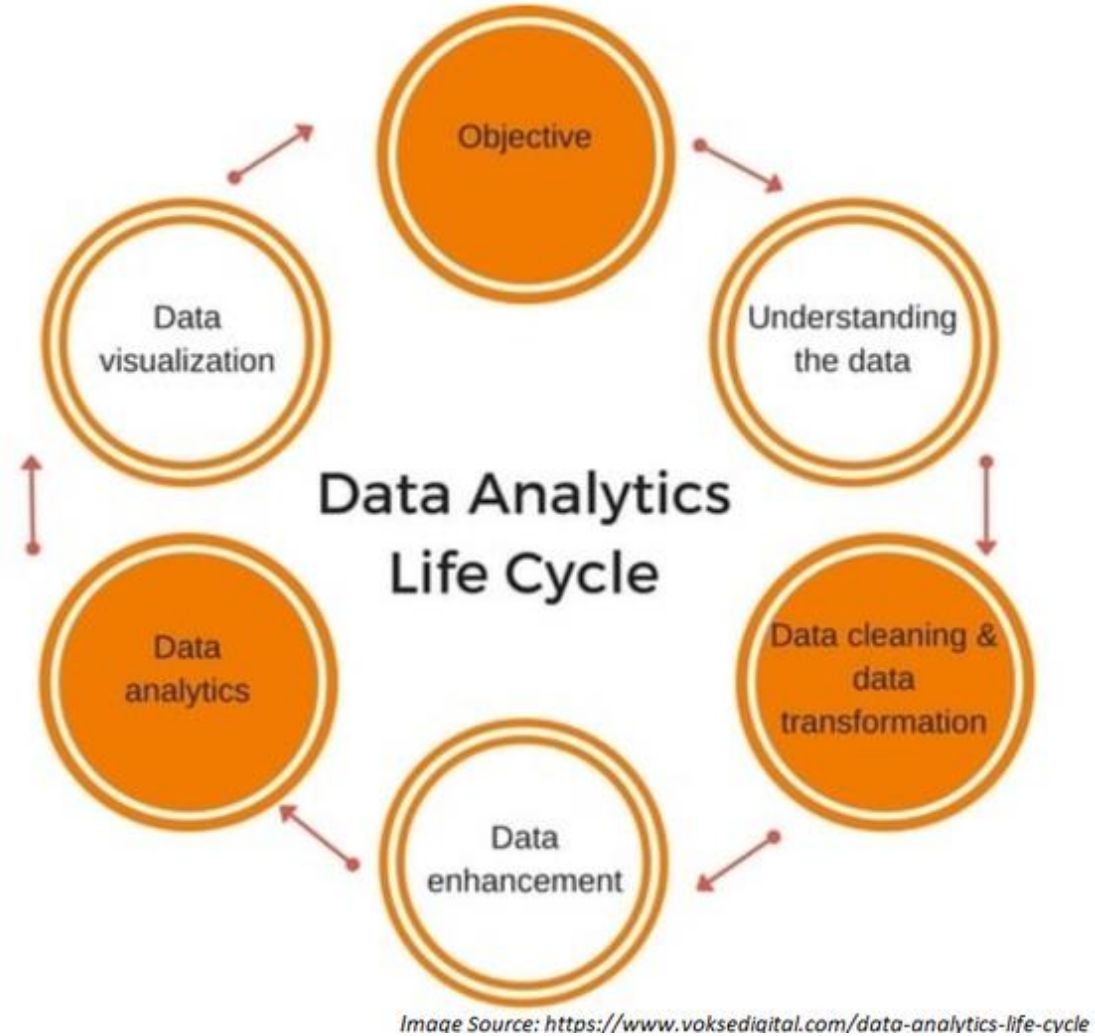
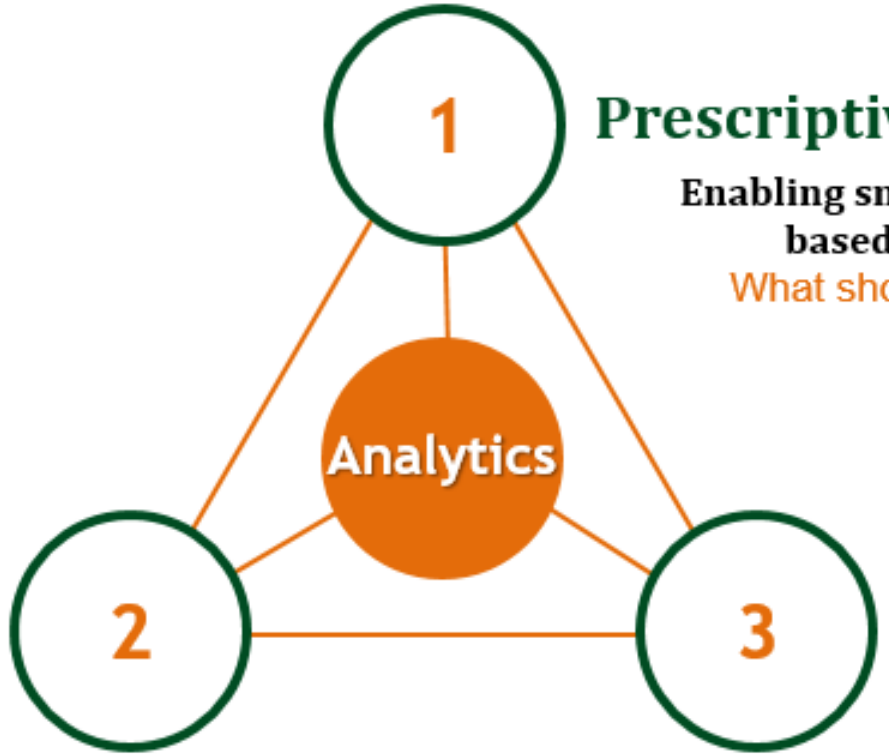


Image Source: https://www.voksediqital.com/data-analytics-life-cycle

Types of Data Analytics

- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics

Types of Data Analysis



Prescriptive Analytics

Enabling smart decisions based on data

What should we do?

Predictive analytics

Predicting the future based on historical patterns

What could happen?

Descriptive analytics

Mining data to provide business insights

What has happened?

Types of Data Analysis

Descriptive

- Aims to help uncover valuable insight from the data being analyzed
- Answers the question **"What has happened?"**

Predictive

- Helps forecast behavior of people and markets
- Answers the question **"What could happen?"**

Prescriptive

- Suggests conclusions or actions that may be taken based on the analysis
- Answers the question
- **"What should be done?"**
- **What Should we do**

Types of Analytics

How does Netflix frequently recommend just the right movie?

Descriptive Analytics
insight into the past

How do grocery cashiers know to hand you coupons you might actually use?

Predictive Analytics
understanding the future

Why do airline prices change every hour?

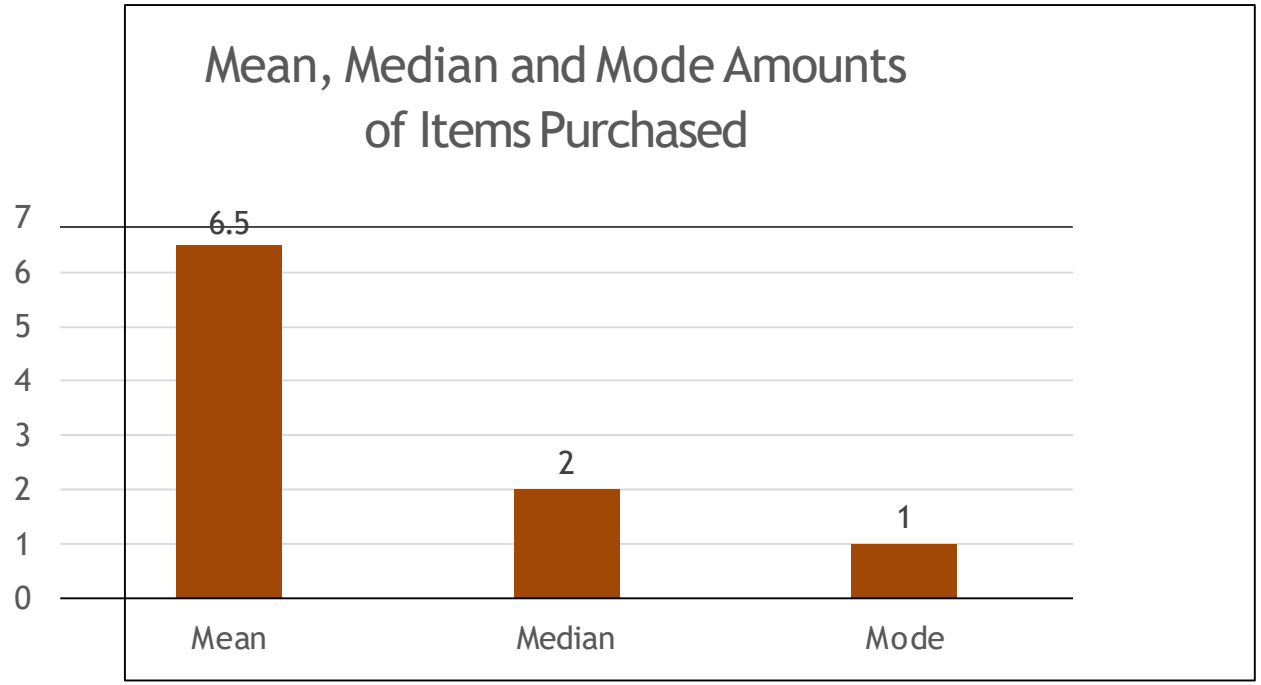
Prescriptive Analytics
advice on possible outcomes

Descriptive

Data Analytics

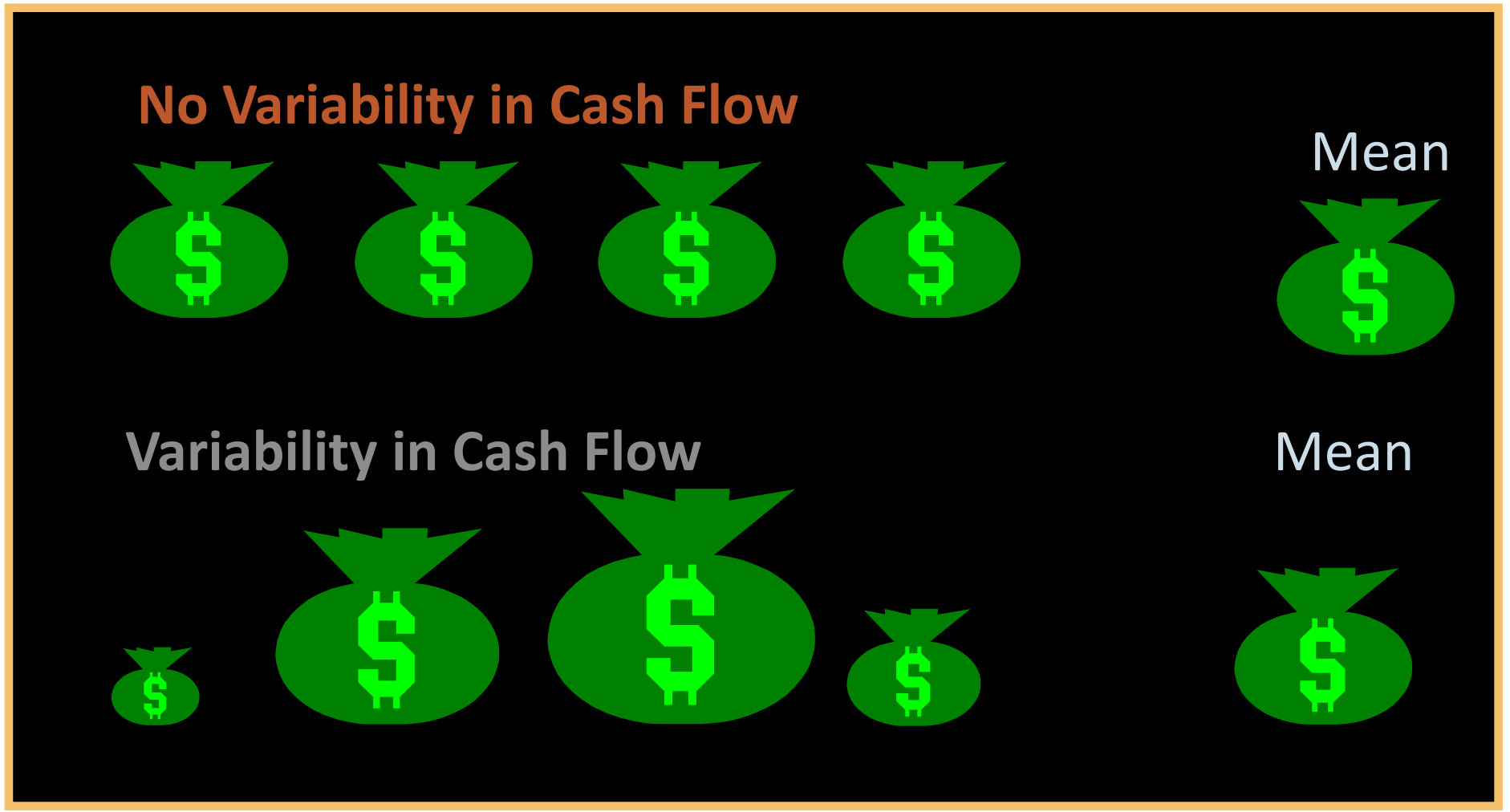
- Though the most simple type, it is used most often.
- Two types of descriptive analysis:

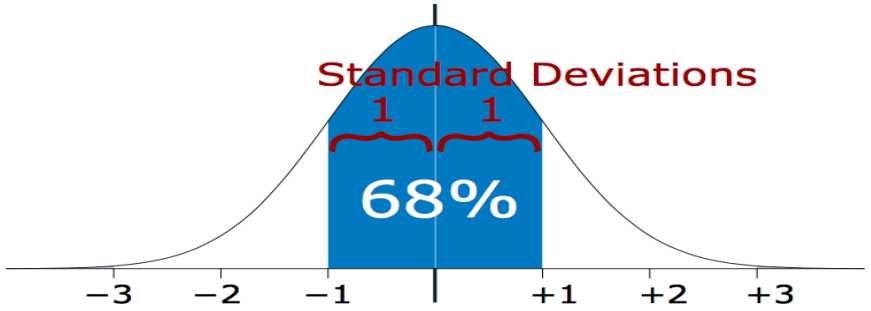
1. Measures of central tendency (tells us about the middle)
   - Mean – the average
   - Median – the midpoint of the responses
   - Mode – the response with the highest frequency
2. Measures of dispersion
   - Range – the min, the max and the
   - distance between the two
   - Variance – the average degree to which each of the points differ from the mean
   - Standard Deviation – the most common/standard way of expressing the spread of data



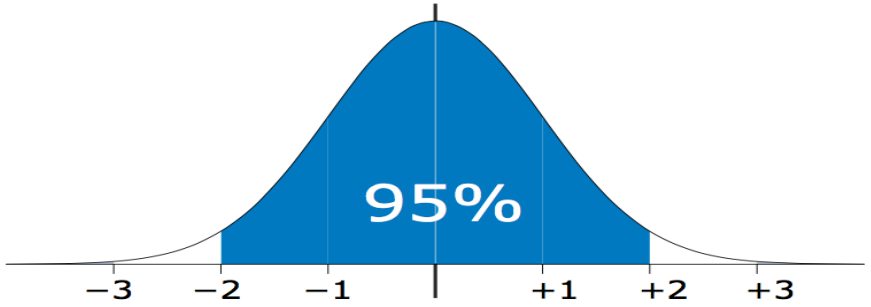
Customer_ID	Items Purchased	Amount Spent
29304	1	1.09
28308	3	44.43
19962	21	218.58
30281	1	73.02

Variability

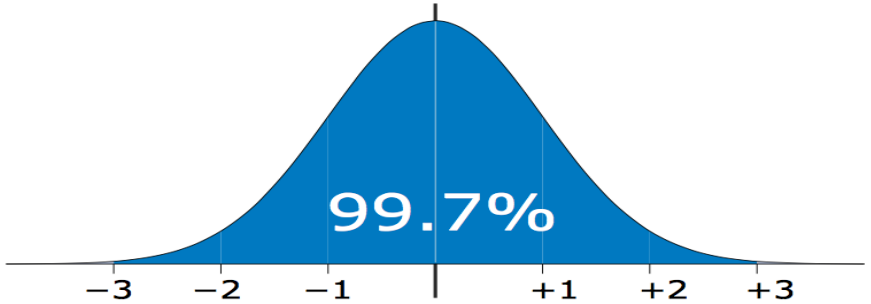




68% of values are within
1 standard deviation of the mean



95% of values are within
2 standard deviations of the mean



99.7% of values are within
3 standard deviations of the mean

Predictive Data Analytics

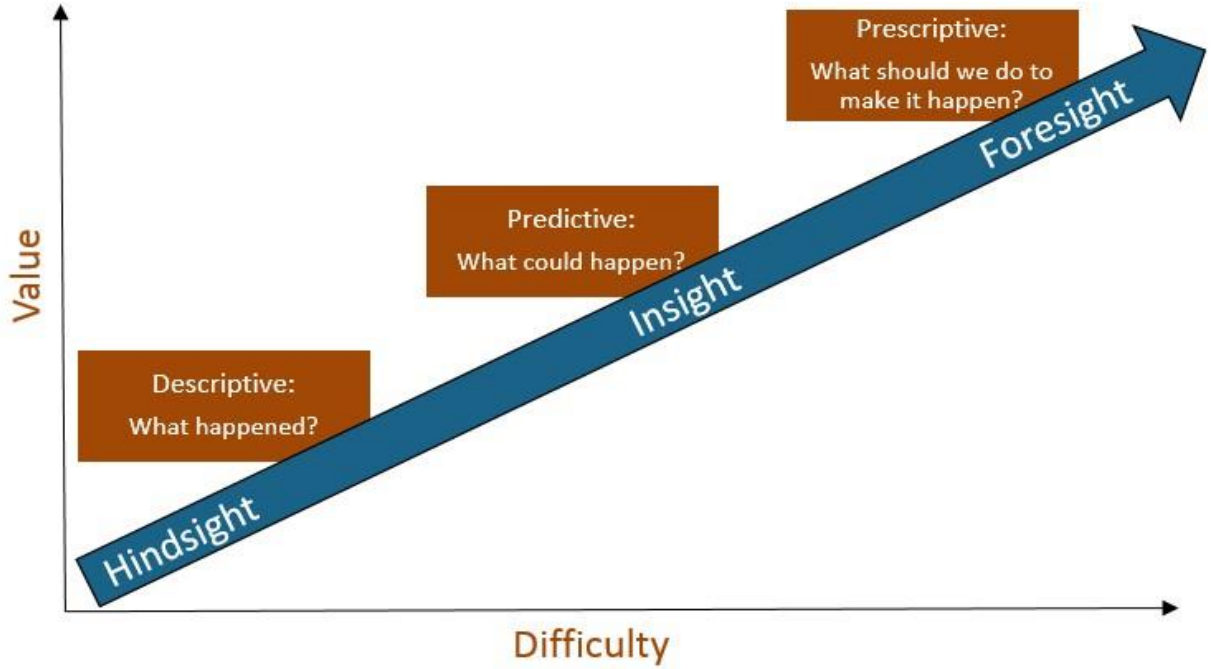
- Some mistake predictive analysis to have exclusive relevance to predicting *future* events.
  - However, in cases such as sentiment analysis, existing data (e.g., the text of a tweet) is used to predict non-existent data (whether the tweet is positive or negative).
- Several of the models that can be used for predictive analysis are:
  - Forecasting
  - Simulation
  - Regression
  - Classification
  - Clustering

Prescriptive Data Analytics

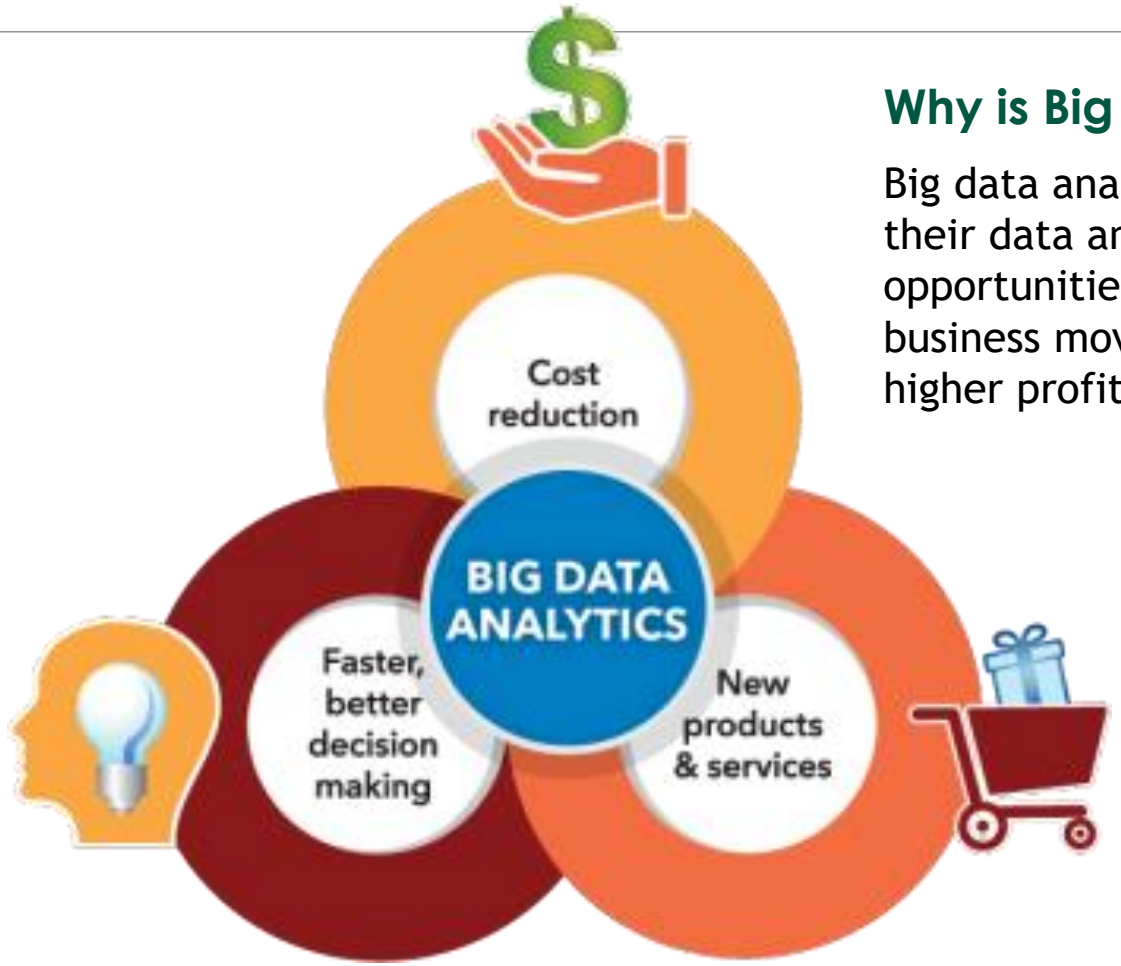
- Decisions can be formulated from descriptive and predictive analysis
  - If I need to cut a product and I know that product C is least preferred and least profitable, I will cut product C.
- However, prescriptive analytics explicitly tell you the decisions that should be made. This can be done using a variety of techniques:
  - Linear programming
  - Integer programming
  - Mixed integer programming
  - Nonlinear programming

Comparing the Three Types of Data Analytics

- Descriptive analysis is most common.
  - Best practice to perform descriptive analyses prior to prescriptive/predictive
    - Understand that distribution, variance, skew, etc., may exclude certain models
- How to know which type of analysis to pursue:
  - How much time do you have?
  - What resources are available to you?
  - How accurate is your data? How accurate do you need the model/analysis to be?
  - How popular/accepted is the model you are considering?
    - Don't subscribe to "that's how we've always done it," but remember to use a model that stakeholders will accept.



Why Big Data Analytics?



Why is Big Data Analytics important?

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers.

Data Analytics Tools

HEAP

Data Analyst/Journalist

Research and extract information from various sources to describe the main features of a data set in support of business operations and decision making.

Data Scientist

Cleanse data, apply various techniques to large data sets and create visualizations to understand to answer more abstract business questions, provide more predictive insights and anticipatory behaviors.

Enter Data Scientists

Data Scientist:

The Sexiest Job In The 21st Century

Harvard Business Review, Oct 2012

A Business analyst is not able to discover insights from huge sets of data of different domains.

Data scientists can work in co-ordination with different verticals of an organization and find useful patterns/insights for a company to make tangible business decisions.

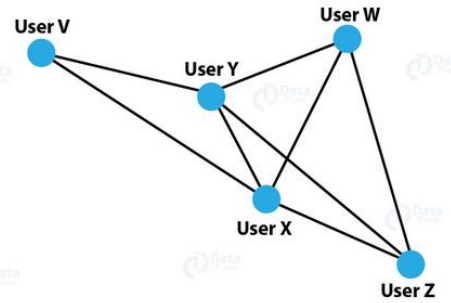
15,000%

INCREASE IN JOB POSTINGS FOR DATA SCIENTISTS IN THE US BETWEEN 2011-12

USE CASES


facebook
Social Analytics
amazon
Improving E-commerce experience
UBER
Optimizing Rides
Bank of America
Increasing Customer experience
Airbnb
Improving Searches
Spotify
Music Recommendation

Facebook

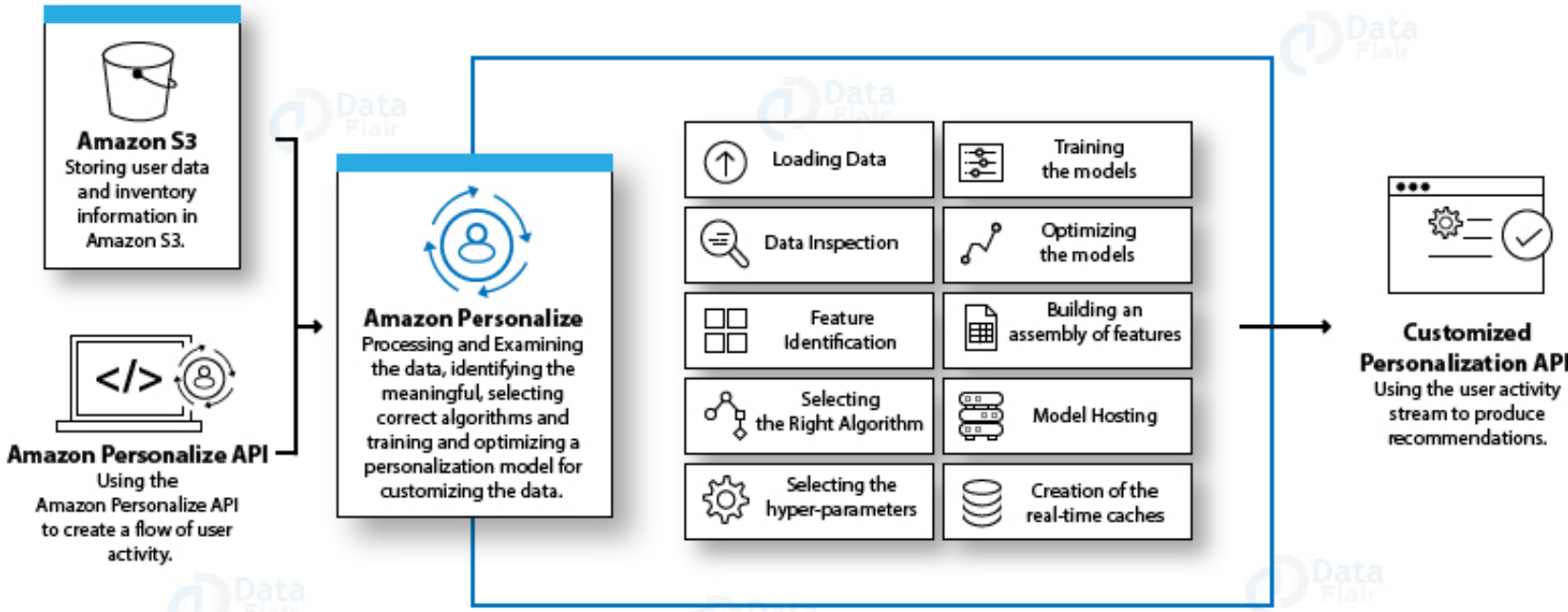


- 2.5 billion monthly active users
- Deep learning: Facebook makes use of facial recognition and text analysis
- Facial recognition, Facebook uses powerful neural networks to classify faces in the photographs
- Uses its own text understanding engine called "**DeepText**" to understand user sentences.
- For targeted advertising: Deep Learning
  - It uses the insights gained from the data to cluster users based on their preferences and provides them with the advertisements that appeal to them.

Amazon

- Amazon heavily relies on **predictive analytics** to increase customer satisfaction.
- Personalized recommendation system.
- This also comes through the suggestions that are drawn from the other users who use similar products or provide similar ratings.
- This also comes through the suggestions that are drawn from the other users who use similar products or provide similar ratings.

Amazon's Recommendation Engine



Amazon

- Fraud Detection
- Amazon has its own novel ways and algorithms to detect fraud sellers and fraudulent purchases.

Uber

- Uber is a popular smartphone application that allows you to book a cab.
- Uber makes extensive use of Big Data.
- Uber has to maintain a large database of drivers, customers, and several other records.
- Whenever you hail for a cab, Uber matches your profile with the most suitable driver.

- **Base (or initial) fare –** A flat fee charged at the beginning of every ride
- **Cost per minute –** How much you are charged for each minute you are inside the ride
- **Cost per mile –** How much you are charged for each mile of the ride
- **Booking Fee (Formerly 'Safe Rides Fee')** – A flat fee to cover Uber's 'operating costs' (Not included for Uber's more luxury services like UberBlack or UberSUV)

Here's how Uber uses the 4 main criteria above to calculate your fare:

*Base Fare + (Cost per minute * time in ride) + (Cost per mile * ride distance) + Booking Fee = Your Fare*

Bank of America - Using Data to Leverage Customer Experience

- BoA are making use of Data Science and predictive analytics
- Banking industries are able to detect frauds in payments and customer information
- Prevents frauds regarding insurances, credit cards, and accounting.
- Banks employ data scientists to use their quantitative knowledge where they apply algorithms like association, clustering, forecasting, and classification.
- Risk modeling - Using Machine Learning, banks are able to minimize risk modeling.

Bank of America - Using Data to Leverage Customer Experience

- Understanding their customers through an intelligent customer segmentation approach: high-value and low-value segments.
- Clustering, logistic regression, decision trees to help the banks to understand the Customer Lifetime Value (CLV) and take group them in the appropriate segments.

Airbnb

- host accommodations as well as find them through its mobile app and website
- Contain massive big data of customer and host information, homestays and lodge records, as well as website traffic
- Ex : In 2014, Airbnb found out that users from certain countries would click the neighborhood link, browse the page and photos and not make any booking.
- In order to mitigate this issue, Airbnb released a different version for the users from those countries and replaced neighborhood links with the top travel destinations. This saw a 10% improvement in the lift rate for those users.

Spotify

- online music streaming giant
- With over 100 million users, Spotify deals with a massive amount of big data
- uses the 600 GBs of daily data generated by the users to build its algorithms to boost user experience.
- In the year 2017, Spotify used data science to gain insights about which universities had the highest percentage of party playlists and which ones spent the most time on it.