greatlearning
Learning for Life

DATA SCIENCE USING PYTHON
Week 2

greatlearning
Learning for Life

GROUND RULES

- Come prepared for these sessions by watching the videos.
  - Concepts will be covered in the videos.
  - Hands-On Application will be covered in Mentor Sessions.

- Submit all assignments on time.

- Let’s be punctual & respect each other’s time.

greatlearning
Learning for Life

DSBA CURRICULUM DESIGN

FOUNDATIONS

Data Science Using Python (2/3)

Statistical Methods for Decision Making

CORE COURSES

Advanced Statistics

Data Mining

Predictive Modelling

Machine Learning

Time Series Forecasting

Data Visualization

DOMAIN APPLICATIONS

Financial Risk Analytics

Web & Social Media Analytics

Marketing Retail Analytics

greatlearning
Learning for Life

LEARNING OBJECTIVE OF THIS MODULE

- Basic Working proficiency in Python
- Basic Data-Manipulation using Python
- Basic Data-Visualization using Python

greatlearning
Learning for Life

LEARNING OBJECTIVES OF THIS SESSION

- Let’s test our understanding

- Hands-On Case Study

- Industry Applications of EDA

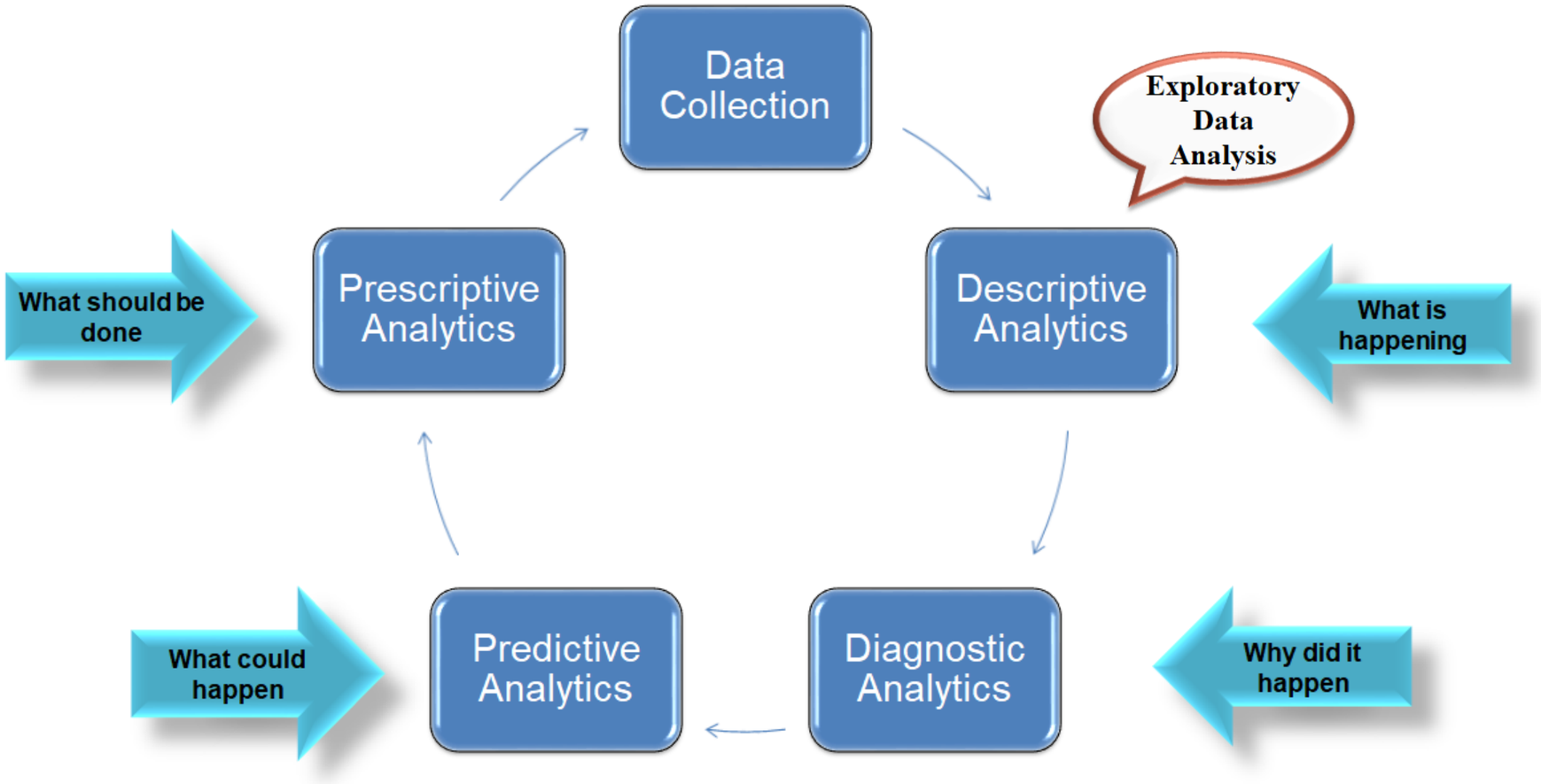
greatlearning
Learning for Life

Let’s test our understanding

- Name few libraries used for visualization?
- What is static and dynamic visualization?
- What are the different types of Visualization techniques?

greatlearning
Learning for Life

Data Science Project Lifecycle



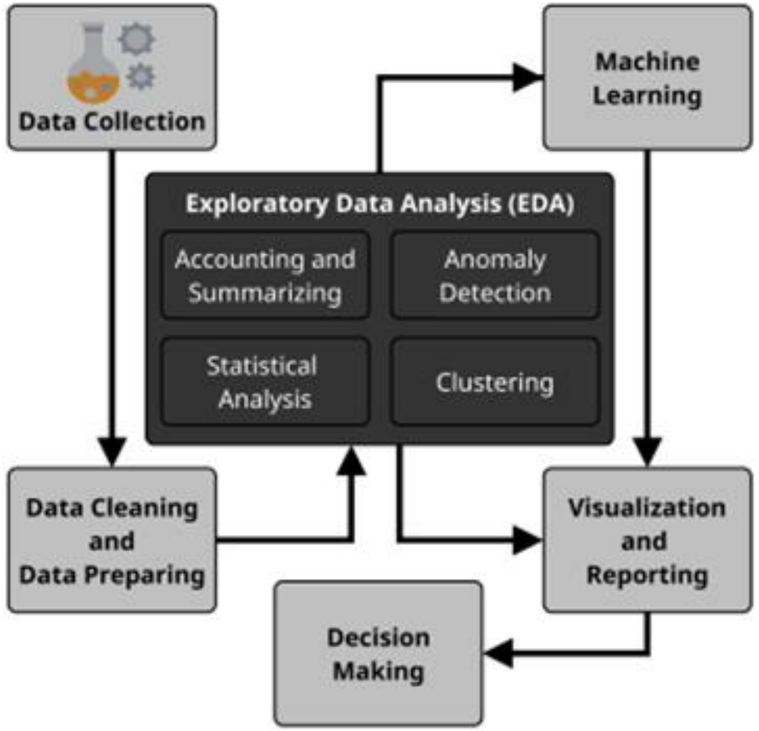
greatlearning
Learning for Life

Exploratory Data Analysis

- Understand the spread of variables in the dataset
- Detect any outliers in the dataset
- Obtain cues on relationship between variables in a dataset
- Spot missing values in the dataset

greatlearning
Learning for Life

How Netflix used Exploratory Data Analysis



A great example of this is the immensely popular show **House of Cards** and how **Netflix** used data to create the show. They started by collecting massive amounts of data from their users, such as the kinds of shows and movies they watched and what actors and directors were the most popular. After crunching the numbers, they were able to identify that director David Fincher was among the most popular, as was actor Kevin Spacey. Interestingly, they also found that the original British version of House of Cards was very well received by these same users. The numbers weighed very heavily in favour of combining these three elements (the show, Spacey and Fincher), which prompted them to buy the rights for the show and create an American version, which released in early 2013.

https://analyticstraining.com/all-you-need-to-know-about-analytics/

Did all this data work? The answer to that is a resounding yes! The show proved to be a massive hit and Netflix gained around 2 million new subscribers in the first quarter of 2013 in the US alone. They also got a further 1 million new users from other markets around the world and this influx of new subscribers helped reimburse the production costs of the show.

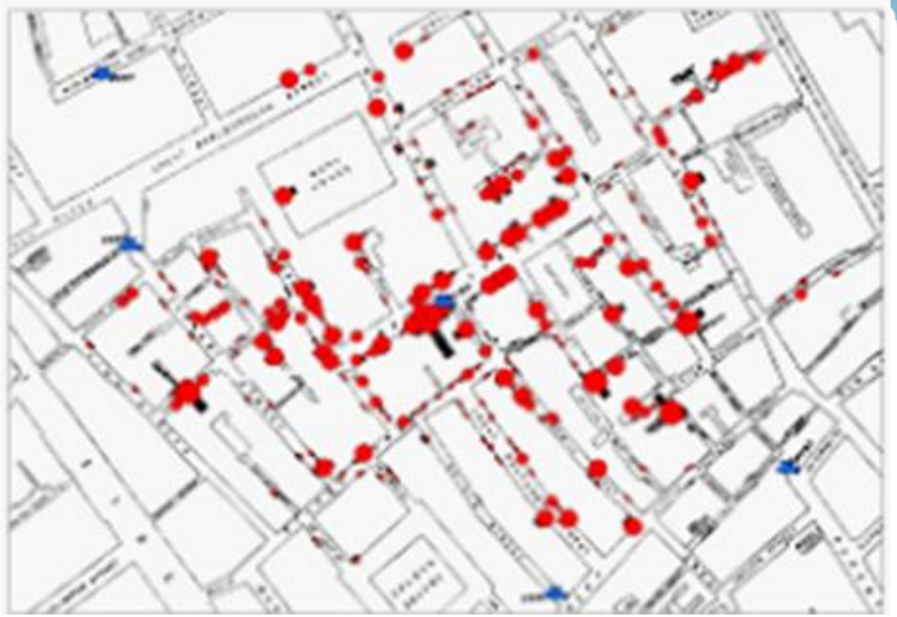
greatlearning
Learning for Life

How Visual Analysis helped cure Cholera

In 1854 there was a major outbreak of **cholera in London**. Many then-prevailing theories failed to associate the transmission of cholera to hygiene, assuming that cholera was non-contagious or that the disease agent was airborne (Vinten-Johansen et al., 2003).

A doctor named John Snow discovered the cause of the outbreak by making a map indicating where the victims lived. The map showed a cluster in Soho, and Snow identified a water pump in Broad Street as the epicenter of the outbreak. Excavations showed that a nearby cesspit leaked fecal bacteria into the pump-well.

Snow convinced officials to remove the handle from the pump, thus ending the outbreak (some accounts even claim that Snow removed the handle himself)



https://www.researchgate.net/publication/228410393_Principles_of_Exploratory_Data_Analysis_in_Problem_Solving_What_Can_We_Learn_from_a_Well-Known_Case

greatlearning
Learning for Life

CASE STUDY - Honey Production

In 2006, global concern was raised over the rapid decline in the honeybee population, an integral component to American honey agriculture. Large numbers of hives were lost to Colony Collapse Disorder, a phenomenon of disappearing worker bees causing the remaining hive colony to collapse. Twelve years later, some industries are observing recovery but the American honey industry is still largely struggling. Now, honey mostly comes from overseas, with 350 of the 400 million pounds of honey consumed every year originating from imports. This dataset provides insight into honey production supply and demand in America by state from 1998 to 2012

greatlearning
Learning for Life

BY THE ALUMs

One of the core elements that has enhanced my learning experience and my successful program completion is watching the videos diligently every week. I could make the most out of different platforms of support like mentoring sessions only because I did my homework well and had my doubts ready to ask my mentor. Video content is truly the heart of the learning process

Aman Arora

NAME HERE
Company name here

greatlearning
Learning for Life

HAPPY LEARNING