LIFE CYCLE OF A DATA SCIENCE PROJECT

STANDARD 8 STEPS

1. BUSINESS UNDERSTANDING
2. DATA ACQUISITION
3. DATA PREPARATION
4. EXPLORING AND MODELLING
5. EVALUATION AND INTERPRETATION
6. DEPLOYMENT
7. OPERATIONS
8. OPTIMIZATION

BUSINESS UNDERSTANDING

- HOW CAN WE TRANSLATE DATA INTO DOLLARS?
- WHAT IMPACT DO I WANT TO MAKE WITH THIS DATA?
- WHAT BUSINESS VALUE DOES OUR MODEL BRING TO THE TABLE?
- WHAT WILL SAVE US LOTS OF MONEY?
- WHAT CAN BE DONE TO MAKE OUR BUSINESS RUN MORE EFFICIENTLY ?

DATA ACQUISITION

- **FLAT FILES**

  Eg: CSV file

- **RELATIONAL DATABASES**

  Microsoft SQL Server, Oracle Database, MySQL and IBM DB2.

- **DATAWAREHOUSE**

  Teradata, Amazon redshift, Oracle and Cloudera

- **TRANSACTIONAL DATABASES**

  Application: Banking, Distributed systems, Object databases, et

- **MULTIMEDIA DATABASES**

  Application: Digital libraries, video-on demand, news-on demand, musical database, etc

- **SPATIAL DATABASE**

  Application: Maps, Global positioning, etc.

- **TIME-SERIES DATABASES**

  Time series databases contains stock exchange data and user logged activities

- **WEBSITE /API**

  Application: Online shopping, Job search etc.

DATA PREPARATION/ CLEANING YOUR DATA

- HANDLING MISSING DATA
- CORRECTING INVALID VALUES
- REMOVING DUPLICATES
- STRUCTURING THE DATA TO BE FED INTO AN ALGORITHM
- FEATURE ENGINEERING

EXPLORING AND MODELLING

- FIND PATTERNS IN YOUR DATA THROUGH VISUALIZATIONS AND CHARTS
- EXTRACT FEATURES BY USING STATISTICS TO IDENTIFY AND TEST SIGNIFICANT VARIABLES
- IN-DEPTH ANALYTICS: CREATE PREDICTIVE MODELS/ALGORITHMS

EVALUATION AND INTERPRETATION

- EVALUATE AND REFINE THE MODEL
- REPORT THE RESULTS

DEPLOYMENT

- API
- MOBILE APP
- WEB APP
- IOT
  - Arduino

OPERATIONS/MAINTENANCE

- DEVELOPING A PLAN FOR MONITORING AND MAINTAINING THE DATA SCIENCE PROJECT IN THE LONG RUN.

OPTIMIZATION

- RETRAINING THE MACHINE LEARNING MODEL IN PRODUCTION WHENEVER THERE ARE NEW DATA SOURCES COMING