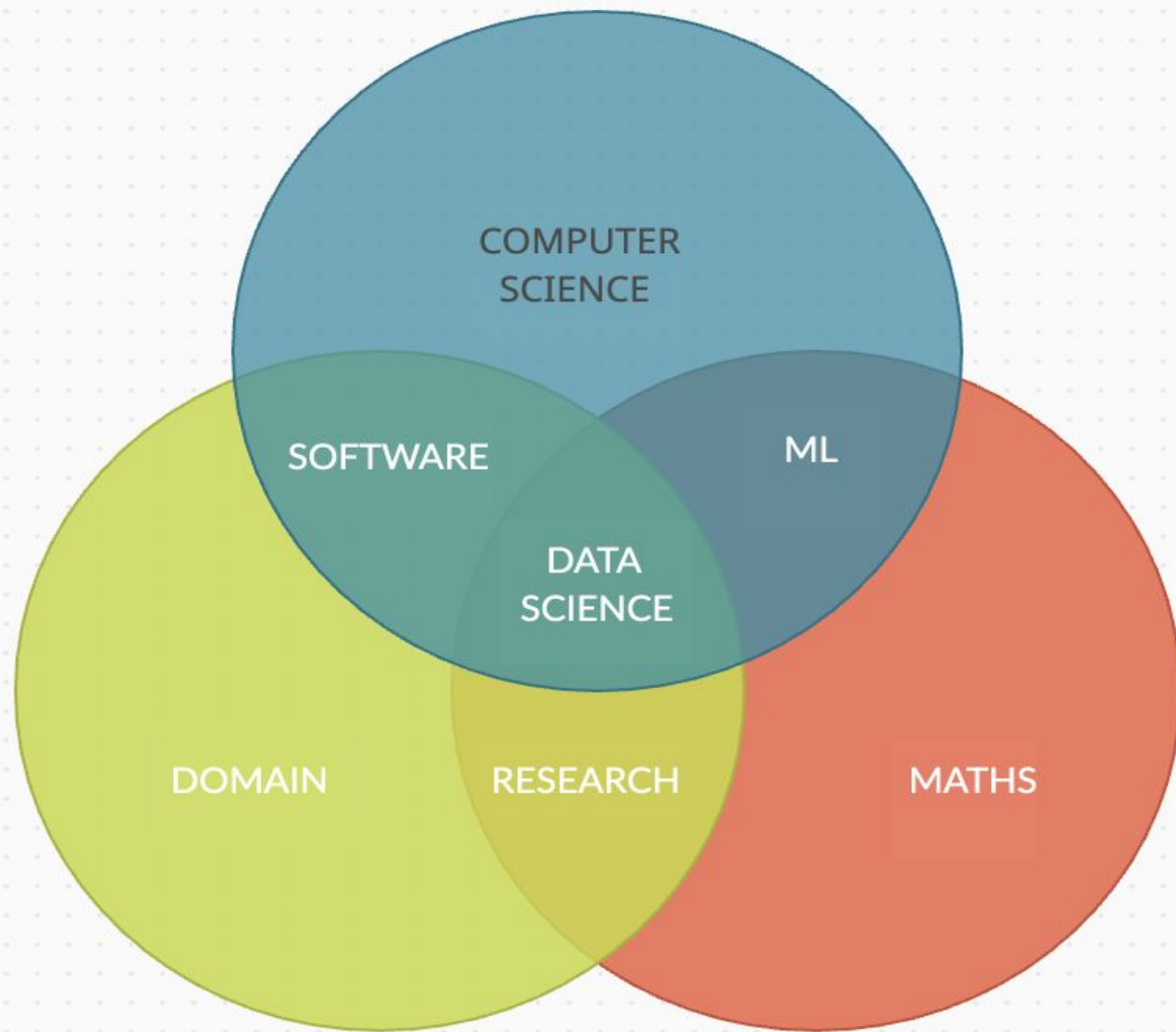


# Basics of Data Science

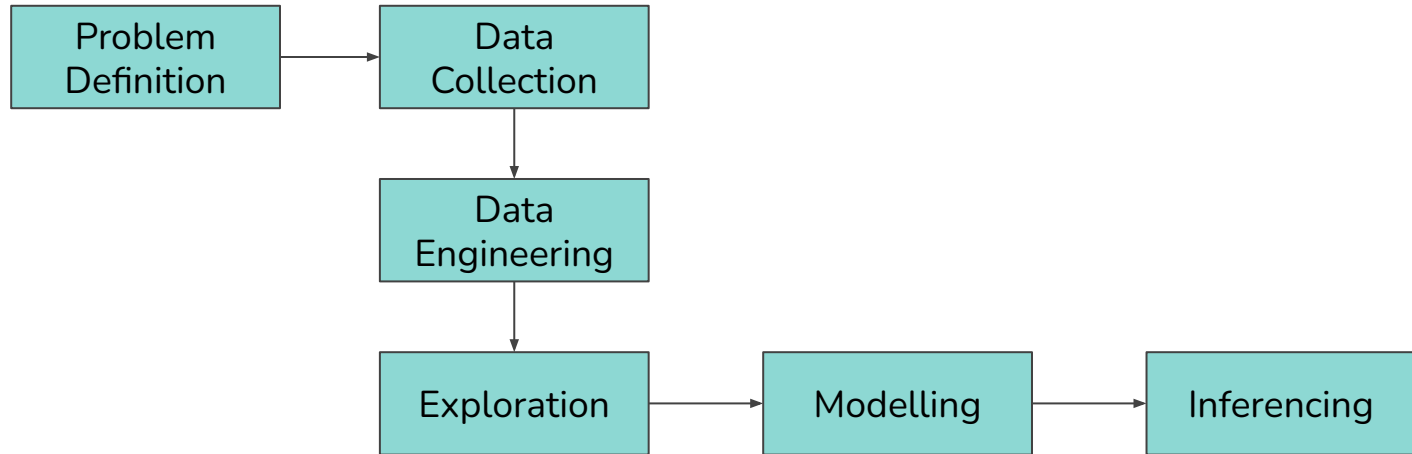
SkillVertex (Satya Pattnaik)







# Data Science Life Cycle





# Various Roles

Data Engineering	Data Analysis	Data Science
Data Pipelines at Scale for DS & DA	Descriptive Analytics	Descriptive, Predictive and Prescriptive
Engineering Heavy	Analysis	Analysis + Science + Engineering
Hadoop, Kafka, Python, Java, Spark	Tableau, Microsoft Power BI	Python, Scikit Learn, Pytorch



# Descriptive vs Predictive vs Prescriptive

Descriptive	Predictive	Prescriptive
What has Happened?	What will Happen?	What if?
Sales till now	Sales in Future	How to Price
Excel, Tableau, Descriptive Statistics	Statistical Modelling, Machine Learning	What if Tools, Optimization



# Data Collection/Cleaning

Identify Data Pertaining to problem

Work with Data Engineers to create correct data pipelines

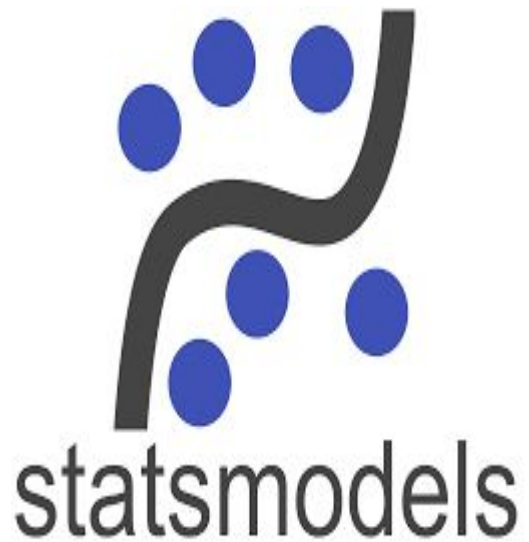
Process the Data

Clean the Data

Explore the Data



## Model

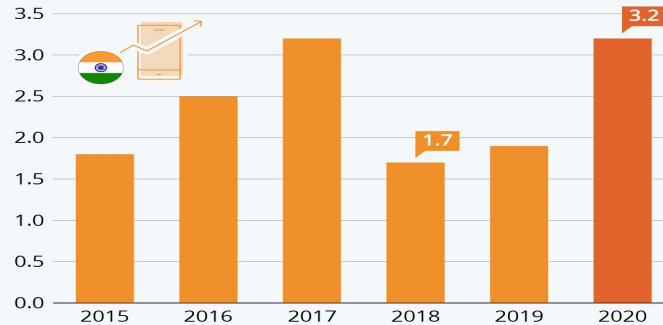




# Descriptive

## Demand for iPhones up in India

Number of iPhones shipped in India by Apple (in millions)



Source: Counterpoint Research

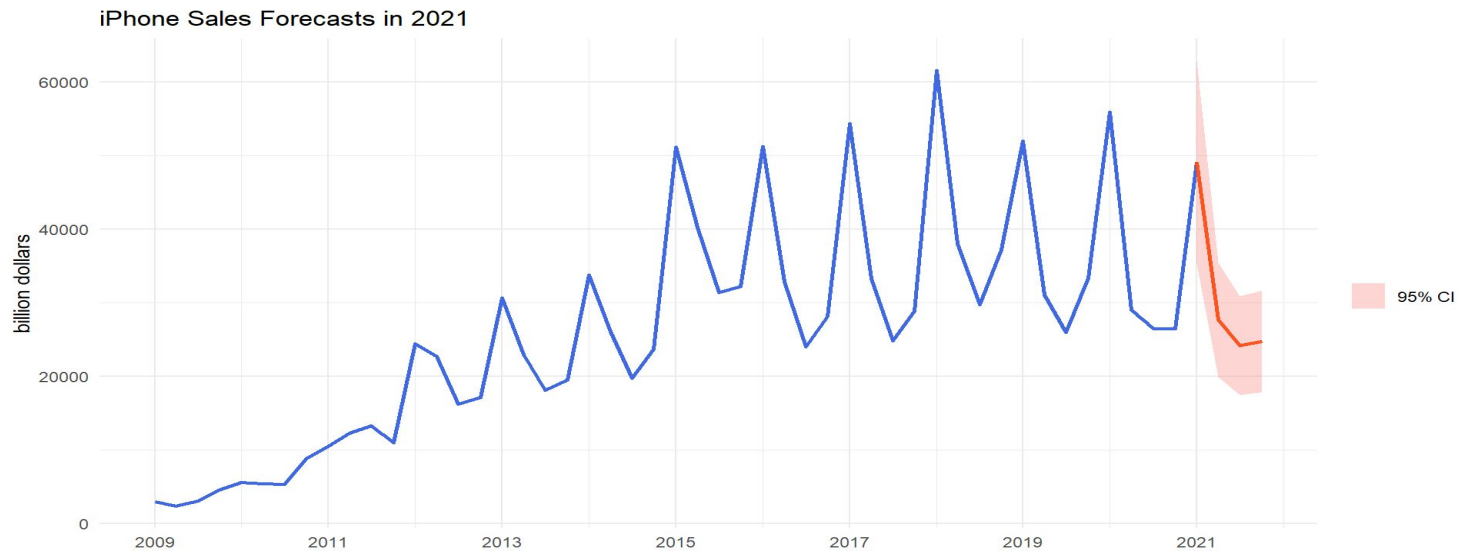


statista





# Predictive





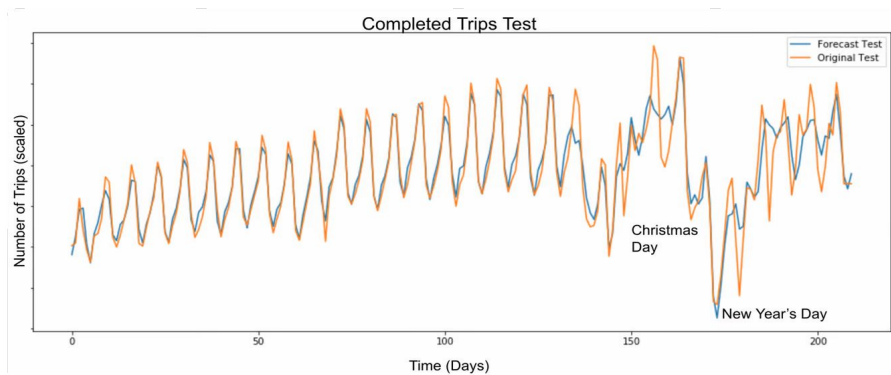
# Real World Problems

- Best Route for Delivery
- Best Movie Recommendations
- Healthcare to Specify Treatments
- Credit Risk Models
- Friends on Social Media



# Real World problems in Data Science

Uber Marketplace Forecasting: A critical element of our platform, marketplace forecasting enables us to predict user supply and demand in a spatio-temporal fine granular fashion to direct driver-partners to high demand areas before they arise, thereby increasing their trip count and earnings.





# Real World problems in Data Science

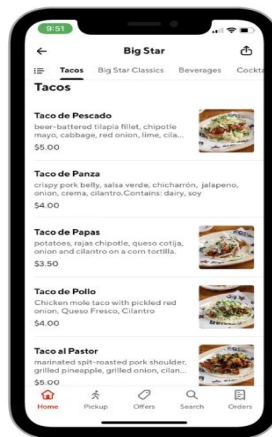
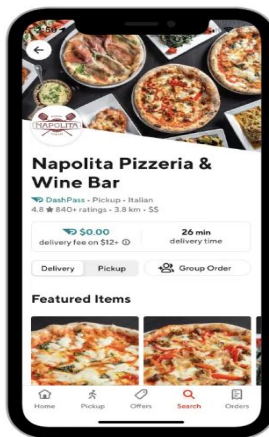
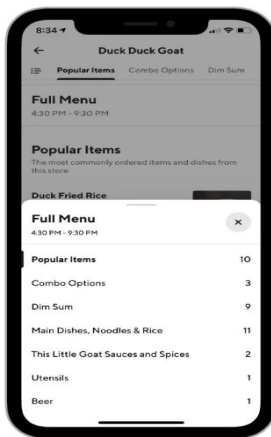
Airbnb Lifetime Value Modelling

id_listing	host_location_city	avg-nightly-price	availability_next_180_days	1_year_revenue
1	London	\$120	50 nights	\$
2	San Francisco	NULL	150 nights	\$\$
3	Tokyo	\$55	NULL	\$\$\$
4	New York	\$100	90 nights	\$\$\$\$



# Real World Data Science Problems

## Doordash Best Menu Selection





# References

1. <https://www.statista.com/chart/24036/iphone-shipments-india/>
2. <https://rpubs.com/crystalwanyulee/698729>
3. <https://www.uber.com/en-IN/blog/forecasting-introduction/>
4. <https://medium.com/airbnb-engineering/using-machine-learning-to-predict-value-of-homes-on-airbnb-9272d3d4739d>
5. <https://doordash.engineering/2020/11/10/uncovering-online-delivery-menu-best-practices-with-machine-learning/>

# Introduction to Machine Learning

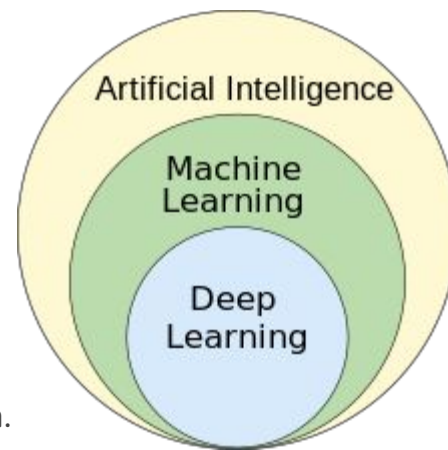
SkillVertex (Satya Pattnaik)





# Definition of ML

- Coined in 1959 by an IBM Employee(Arthur Samuel)
- Study of algorithms which can learn relationships from data.
- Sub Field of Artificial Intelligence.
- Formal Definition:[Tom M. Mitchell](#) provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$  if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ ."







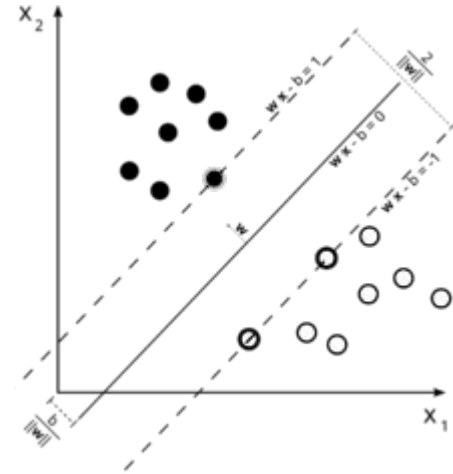
# Approaches

- Supervised ML
- Unsupervised ML
- Dimensionality Reduction
- Reinforcement Learning



# Supervised Machine Learning

- Learns Linear/non-Linear Mapping  
 $f(X) \rightarrow Y$
- Training Data
- Objective Function
- E.g:
  - Support Vector Machine
  - Decision Tree
  - Neural Networks
  - Logistic Regression





# Unsupervised Machine Learning

- Major Types:
  - Clustering
  - Association
- No Categories/labels reqd.
- Divides Data Into Classes.
- Grouping based on Similarities and Differences.



# Differences

	Supervised	Unsupervised
Input	Labelled	Not labelled
Approach	Learn relationship between X(Features) & Y(Labels)	Divide Data into Classes
Application	Regression/Classification	Clustering/Association



# References

1. [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)