|  |  |
| --- | --- |
| 1. (a) Sample Collection: How to collect the relevant Data?  Does the data Match the Problem?  Does the data Represent Reality?  How to effectively Sample real Data? | 1. (b) Practical Motivation: What is the Real-Life Problem?  Can you relate the Problem to Data?  Would Data help you in Practice?  How to Identify a Data Science case? |
| "How to find best suited listing for Users based on preferences"  - listing\_id based on price  - type of accomadation (space)  - reviews on certain attributes (eg. Cozy, quiet, etc)  - day of the week | |
| 2. (a) Data Preparation: How to Prepare the relevant Data?  Is the data Clean enough to Analyze?  Is the data Structured for Analysis?  How to prepare Raw Data for Analysis? | 2. (b) Problem Formulation: What is the Data Science Problem?  How do you Formulate it using Data?  How do you Solve it using the Data?  How to intelligently Construct a Problem? |
| Structured-Highly Org, Easy to Analyze/Mine, Numeric/Factor(Excel), Time Series(Stocks), Network(Map)  Unstructured-Highly Unorg, Text(Web), Image(Web), Voice(Song), Vids(Youtube) | (Prediction)Numeric-Regression > Linear Regression, Tree Models, Neural Network  (Prediction)Classes-Classification > Logistic Regression, Tree Models, Neural Network  (Detection)Structure-Clustering > Distance: Euclidean;Jaccard, K-Algorithm, Hierarchical Model  (Detection)Anomaly-Anomaly Detection > Cluster-Analysis Detection, Nearest Neighbour Model, Support Vector Model  (Decision)Action-Adaptive Learning > Reinforcement Learning, Monte-Carlo, State-Action-Reward, Q-Learning, Deep Reinforcement |
| 3. (a) Exploratory Analysis: How to Explore the acquired Data?  How to effectively Mine the Data?  How to Compute the vital Statistics?  How to gain basic Insight from Data? | 3. (b) Statistical Description: How to clearly Describe the Data?  How do you Summarize the Data?  Which vital Statistics are relevant?  How to succinctly Represent the Data? |
| Mean/Median/SD/Summary | |
| 4. (a) Analytic Visualization: How to clearly Visualize the Data?  How to visually Represent Statistics?  How to highlight “Interesting” Traits?  How to represent the Data for the Humans? | 4. (b) Pattern Recognition: How to Identify structure in Data?  Can you See the known Patterns?  Can you Discover unknown Traits?  How to find Intrinsic insight from the Data? |
| Box-Plot, Histogram, Density Plot, Violin Plot  Bi-Variate - Joint-Plot Relation (Correlation Coefficient, Correlation Matrix and Plot)  Multi-Variate - Pair-Plot (Historgram, Scatter Plot) | Pattern Recognition: Distribution of Variables, Inter-Variable Dependence |
| 5. (a) Algorithmic Optimization: How to form “Learning” algorithms?  How to Reduce Errors in Learning?  How to Generalize the Algorithms?  How to optimally Learn from the Data? | 5. (b) Machine Learning: How to Learn from the Data?  Can you formulate the “Learning”?  Can you automate the “Learning”?  How to efficiently Learn from the Data? |
| Regression: Linear Model, Goodness of Fit-Mean Square Error(MSE)low-good, Explained Variance(R^2) high-good  Classification: Decision Tree(GINI), Goodness of Fit-Classification Accuracy(TP/TN)  Clustering: K-Means (Centroid Cluster)  Anomaly: Nearest Neighbour (Outliers) | Supervised: Regression, Classification: (Split, Train, Test, Predict)  Unsupervised: Clustering, Anomaly, Reinforcement: (Given data, Find groups, Justify), (Find the difference), (Learn from mistakes) |
| 6. (a) Information Presentation: How to present Analysis Outcomes?  How to present Descriptive Analysis?  How to present Inferential Analysis?  How to Communicate your Data Analysis? | 6. (b) Information Presentation: How to present Analysis Outcomes?  How to present Descriptive Analysis?  How to present Inferential Analysis?  How to Communicate your Data Analysis? |
| 7. (a) Ethical Consideration: Does the Analysis violate Legality?  Does the Decision violate Ethics?  How to Responsibly work in Data Science? | 7. (b) Intelligent Decision: How to take Decisions in Practice?  Can you Decide based on the Data?  Can you Optimize the Outcomes?  How to Solve a real life problem by Data? |
|  |  |
|  |  |