1 Week 1

1. How do you see data ?
   - Good Decisions are based on an accurate understanding of Good data.
   - Present Data in a Precise, Concise and Understandable Way.
   - Two Types of data
     - Categorical
     - Numerical: Discrete and Continuous
   - Core Principle on which visualization of data is done: Nature of data dictates which visualization to use.
2. Benefits of visual representation of data.
   - Communicate complex information concisely and precisely.
   - Create a "picture" for reasoning about and analysing quantitative and conceptual information.
   - Provides "Information Rich View" at a glance.
   - Directs attention towards content rather than methodology.
   - Describe, Explore and Summarize a set of numbers.
   - Convey messages about significance of data.
3. 4 Principles of effective visualization.
   - Know Purpose
   - Ensure Integrity
   - Maximize data ink and Minimize non data ink
   - Show your data, annotate
4. Executing your information display is a 3 step process.
   - Defining the message
     - What am I trying to communicate ?
     - Should I use text, table, graph or a combination ?
     - The message/statistic you want to emphasize
   - Choosing Form
     - What is the message ?
     - What design principles lead to quick cognitive processing & effective communication ?
     - Whether to display the data as a table or a chart
   - Creating Designs
     - How do I make the message clear at a glance ?
     - Avoid 3D effects, Avoid legends(Use Labels), Avoid contrasting borders around objects, Use annotations to highlight key data changes or to focus on specific data points.
5. Dashboard
   - A visual display of the most important information needed to achieve one or more objective that has been consolidated on a single screen so it can be monitored & understood at a glance.
   - Scan the big picture, Zoom in on important specifics, Link to supporting details.

2 Week 2

1. Probability Distributions
   - Trace-Driven Simulation: Data values themselves used directly in simulations.
   - Fit: Use a theoretical distribution for the data.
   - Data values could be used to define empirical distribution.
2. Empirical Distributions
   - Using data we build our own distributions.
   - Define density/Distribution function
   - Estimate Parameters
   - Ungrouped data: $X_1 \leq X_2 \leq X_3 \leq ... \leq X_n$

$$E(x) = \begin{cases} 0 & \text{for } x < X_1 \\ \frac{i-1}{n-1} + \frac{x-X_i}{(n-1)(X_{i+1}-X_i)} & \text{for } X_i \leq x < X_{i+1},\ i = 1, 2, ..., n-1 \\ 1 & \text{for } X_n \leq x \end{cases}$$

   - Grouped Data : $nX_j^{'}$s are grouped in k adjacent intervals so that the jth interval contains nj observations, $n_1 + n_2 + ... + n_k = n$
     Intervals: $(a_0, a_1), (a_1, a_2), ..., (a_{k-1}, a_k)$,
     $G(a_0) = 0, G(a_j) = \frac{n_1+n_2+n_3+...+n_j}{n}$

$$G(x) = \begin{cases} 0 & \text{for } x < a_0 \\ G(a_{j-1}) + \frac{x-a_{j-1}}{a_j-a_{j-1}}[G(a_j) - G(a_{j-1}] & \text{for } a_{j-1} \leq x < a_j,\ j = 1, 2, 3, ..., k \\ 1 & \text{for } a_k \leq x \end{cases}$$

3. Clues from summary statistics
   - Symmetric distributions: mean $\approx$ median, eg: Normal Distribution
   - Coefficient of Variation(cv): Ratio of Standard deviation & mean, $\frac{\sigma}{\mu}$
     Continuous Distributions: cv $\approx$ 1, eg: Exponential Distribution
     Right/Positive skewed histogram: cv $>$ 1, eg: log normal distribution
   - Lexi$^{'}$s ratio: Same as cv for Discrete Distributions.
   - Skewness(v): Measure of symmetry of a distribution
     v = 0, Normal Distribution
     v $>$ 0, right skewed(exponential distribution)
     v $<$ 0, left skewed
4. Parameter Estimation
   - Once distribution is guessed, next step is estimating parameters of the distribution.
   - Most common method used is MLE.
5. Goodness of Fit
   - Can be checked by
     - Frequency Comparison(a bit technical)
     - Probability Comparison(Visual tool)
     - Goodness of Fit test(statistical test for goodness)
   - Quantile-Quantile Plot(Q-Q Plot)
     - Graph of q$_i$ quantile of model vs q$_i$ quantile of sample distribution.
     - $x_{q_i}^M = \hat{F}^{-1}(q_i)$
     - $x_{q_i}^S = \tilde{F}^{-1}(q_i) = x_i, i = 1, 2, 3, ...$
     - If our distribution is correct, then we will get a line with slope 1 and intercept 0 (Linear) & $x_{q_i}^M \approx x_{q_i}^S$
     - Amplifies difference between the tails of model distribution.

- Probability-Probability Plot(P-P Plot)
  - Graph of model Probability $\hat{F}(X_i)$ vs Sample Probability $\tilde{F}_n(X_i)$
  - Valid for both Continuous and Discrete data sets.
  - I chosen distribution is correct then $\hat{F}(X_i) \approx \tilde{F}_n(X_i)$, the plot will be linear with slope 1 and intercept 0.
  - Amplifies differences between middle portion of the model distribution.
- Goodness of fit tests
  - Statistical Hypothesis test that is used to assess formally whether observations are independent samples from a particular distribution.
  - H_0: Observation are independent.
  - Chi-Square Test
    * Require frequency tables: Bins, Object Frequency, Expected frequency.
    * Calculate test statistic $\chi^2 = \frac{\Sigma(O_i - E_i)^2}{E_i}$
    * Compute p-value, if it is less than significant level(α) then reject H_0
    * Compute tabulated $\chi^2_{k-p-1,\alpha}$, if $\chi^2_{tabulated} < \chi^2_{computed}$ then reject H_0.
      p = number of parameters
      k = number of bins

3 Week 3

1. Ordinal Data: Categorical data which can be ordered.
2. Conditional Probability: $\frac{\text{Joint Probability}}{\text{Marginal Probability}}$
3. Conditional Probability can be compared using Joint Probability table(Contingency Table)
4. Baye$'$s Rule
   - Posterior Probability can be found using initial probability and additional information
   - $P(A \cap B) = P(A|B)P(B)$
   - $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$
5. Chi-Squared Test of Independence
6. Null Hypothesis H_0: Categorical Variables are independent.
7. Alternate Hypothesis H_1: Categorical variables are not independent.
8. Example Table:

City\Preferred Brand	Brand A	Brand B	Brand C	Total
Mumbai	279	73	225	577
Chennai	165	47	191	403
Total	444	120	416	980

   - Independent(Explanatory) Variable is City
   - Dependent(Response) Variable is Brand Preference
   - f_o: Observed Frequency(from Samples)
   - f_e: Expected Frequencies, if variables were independent = $\frac{\text{Row Total} \cdot \text{Column Total}}{\text{Overall Total}}$
   - degree of freedom = (number of rows - 1) $\cdot$ (number of columns - 1)

4 Week 4

1. Demand Response Curve
   - Properties: Non-Negative, Continuous & Differentiable and Generally Downwards Slopping
   - Price Sensitivity = $\frac{D(P_2) - D(P_1)}{P_2 - P_1}$
   - Demand Elasticity = $-\frac{\frac{D(P_2) - D(P_1)}{D(P_1)}}{\frac{P_2 - P_1}{P_1}}$

2. Linear Response Curve
   - $D(P) = D_o - mP$
   - Satiating Price $P_s = \frac{D_o}{m}$; $D(P_s) = 0$
   - Demand at $P = 0$ is D_o
   - Elasticity $\varepsilon = \frac{mP}{D_o - mP}$
3. Constant Elasticity Curve
   - $D(P) = cP^{-\varepsilon}$
   - $c =$ Demand when $P = 1$
   - Revenue R $= P \cdot D(P)$
4. Elasticity
   - $\varepsilon < 1$: Inelastic Product Demand, Increase Revenue $\implies$ Increase Price
   - $\varepsilon > 1$: Elastic Product Demand, Increase Revenue $\implies$ Decrease Price
5. Simple Linear Regression can be used to fir Linear curves
   - Loss/Error $= \frac{\Sigma(y-\hat{y})^2}{N}$
   - error term $e = (y - \hat{y})^2 \sim N(0, \sigma_e^2)$
   - error terms are independent, have equal variance and are normally distributed.

5 Week 5

1. Constant Elasticity Model
   - $D(P) = cP^{-\epsilon} \implies \log(D(P)) = \log(c) - \epsilon \log(P)$
   - Transformation based on which quadrant graph lies in
     Ist Quadrant: $(x, y) \longrightarrow (x^2, y^2)$
     IInd Quadrant: $(x, y) \longrightarrow (\log(x), y)$ or $(\frac{1}{x}, y)$
     IIIrd Quadrant: $(x, y) \longrightarrow (\log(x), \log(y))$ or $(\frac{1}{x}, \frac{1}{y})$
     IVth Quadrant: $(x, y) \longrightarrow (x, \log(y))$ or $(x, \frac{1}{y})$
2. Performance Metrics
   - R^2: Amount of variability explained
     Formula: $1 - \frac{sumsquaredregression(SSR)}{totalsumofsquares(SST)}$ where, SSR: $\Sigma(y_i - \hat{y}_i)^2$ and SST: $\Sigma(y_i - \bar{y}_i)^2$
   - Multiple R: Coefficient of Correlation, Strength of Association
     Formula: $\sqrt{R^2}$
   - Std Error: Standard Deviation of Error Terms
   - F Statistic: $\frac{MSE}{MSofRegression}$
   - Confidence Level: Generally 0.95
   - α value: 1 - Confidence Level. If it is greater than p-value then we reject null hypothesis H_0
3. Optimal Pricing
   - Revenue Maximization
     Revenue $= P.D(P)$
     Set $\frac{\delta R(P)}{\delta P} = 0$ and $\frac{\delta^2 R(P)}{\delta P^2} < 0$
   - Profit Maximization
     Profit = Revenue - Cost
     Assuming Marginal Cost = C $\implies \pi(P) = P.D(P) - C.D(P)$

6 Week 6

1. Multiple Linear Regression

   - $y = w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n + \epsilon$, ϵ $N(0, \sigma_\epsilon^2)$
   - Adjusted R^2: Adjusts for the amount of X_i
     Formula: $1 - \frac{(1-R^2)(N-1)}{N-p-1}$, where N is the total sample size and p is the number of independent variables.
   - Calibration Plot: Scatter Plot between y and $\hat{y}$
   - R: Correlation between y and $\hat{y}$
   - Col-linearity: Very high correlation between explanatory variables.
   - Path Diagram: Schematic Drawing of relationship among explanatory variables and response

2. Variance Inflation Factor

   - Quantifies amount of unique variation in each explanatory variable and measures effect of col-linearity
   - VIF of $X_j = \frac{1}{1-R_j^2}$
   - R_j^2 is coefficient of determination in regression of X_j on ALL other explanatory variables.
   - $se(b_1) = \frac{se}{\sqrt{n}}.\frac{1}{s_x}$, with VIF $se(b_1) = \frac{se}{\sqrt{n}}.\frac{\sqrt{VIF(X_1)}}{s_x}$

7 Week 7

1. Logistic Regression(Classification)

   - ODDs(Success in some event) $= \frac{Pr(Y=1)}{Pr(Y=0)}$
   - logit $= \log(ODDs) = w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n$
   - $Pr(Y=1) = \hat{y} = \frac{e^{w_0+w_1x_1+w_2x_2+...}}{1+e^{w_0+w_1x_1+w_2x_2+...}}$
   - We are going to maximize log likelihood or minimize negative log likelihood
     minimize $-(y\log(\hat{y}) + (1-y)\log(1-\hat{y})$
   - if $\hat{y} \geq$ cutoff(generally around 0.5) then prediction is 1 else 0
   - Performance Metrics

	Predicted 1	Predicted 0
Actual 1	True Positive(TP)	False Negative(FN)
Actual 0	False Positive(FP)	True Negative(TN)

     Accuracy: $\frac{TP+TN}{Total}$
     Precision: $\frac{TP}{TP+FP}$
     Sensitivity/Recall: $\frac{TP}{TP+FN}$
     F1 Score: $\frac{2.Precision.Recall}{Precision+Recall}$

8 Week 8

1. Efficiency(Productive Efficiency)

   - Effective utilization of resources for maximization of benefits
   - Productive efficiency is an aspect of economic efficiency focusing on maximizing output under given conditions.
   - Productive efficiency "frontier" are all combination of o/p such that the production of one product cannot be increased without sacrificing the output of the other.

2. Efficiency Measurement

   - In simplest way: $\frac{output}{input}$
   - But there can be multiple types of inputs: Labor, Infrastructure, Money(Resources go as inputs)
   - Outputs can be customers served/acquired, Profits, Sales volume
   - How has the organization performed is the output

3. Optimization Method - Data Envelopment Analysis

    - Non Parametric Mathematical method to find production frontier. Used to calculate productive efficiency of an economic unit. Economic unit is refereed to as "Decision Making Unit". Measures Relative Efficiency, formulates optimization problem for each DMU
    - DEA Logic
        - For multiple inputs and outputs define weighted ratio. Some inputs cannot be directly added, define weight for each input. Similar for outputs.
        - Let each DMU choose input and output weights to its advantage
        - For each DMU:- Objective constraint: Maximize its efficiency by choosing its weights carefully Efficiency $\epsilon[0, 1]$ and using these weights none of the other DMU's should get efficiency > 1 Using these weights is it cannot achieve an efficiency of 1, then this DMU is truly inefficient. Using these weights if a different DMU gets an efficiency of 1, then that DMU is really good.
    - DEA Mathematical Formula
      K = Number of DMU's considered in the data set
      N = Number of inputs considered, M = Number of outputs considered
      I_{ik} = Recorded value of input i for DMU k, O_{jk} = Recorded value of output j for DMU k
      x_{ik} = Weight assigned to input i by DMU k, y_{jk} = Weight assigned to output j by DMU k
      E_k = Efficiency of DMU k = $\frac{WeightedOutput}{WeightedInput} = \frac{\Sigma y_{jk} O_{jk}}{x_{ik} I_{ik}}$
      Subject to $E_k < 1, k = 1, 2, ..., K$. Since, Linear problems are much easier to solve we linearize it.
      For each DMU: Max $\Sigma y_{jk} O_{jk}$ for all j
      Subject to $\Sigma x_{ik} I_{ik} = 1$
      $\Sigma y_{j1} O_{j1} \leq \Sigma x_{i1} I_{i1}$
      $\Sigma y_{j2} O_{j2} \leq \Sigma x_{i2} I_{i2}$
      ...
      $\Sigma y_{jK} O_{jK} \leq \Sigma x_{iK} I_{iK}$

9 Week 9

1. Prescription for inefficient DMU

    - Important Economic Concepts: Types of Efficiencies, and Disposable inputs/outputs
    - The DMU which is NOT on the efficiency frontier should move in both horizontal AND vertical direction.
    - Draw a line from origin to inefficient DMU, then move the DMU along the line towards the efficiency frontier. The point of intersection of this line and efficiency frontier is called Hypothetical Composite Unit(HCU).
    - The efficient DMU's which make up the part of the frontier where HCU resides are the reference DMU's for the inefficient one. In terms of optimization problem, we get the value of dual variable or shadow price, the variables whose value is not 0 correspond to the reference DMU.

10 Week 10 and 11

1. Consumer Choice Model

    - Thinking Process: Consider a consumer comparing four products, we have data for two attributes on these variants, we also have consumer choices data available for us(Consumer provides his/her choices).
    - Marketer would like to know how important each attribute is.

2. Conjoint Analysis

    - It is the analysis of features considered jointly. It constructs a value system and can be used to arrive at the "best" product.
      A Family of techniques that model choice by decomposing overall preference in terms of relative values of components or attributes to respondents.
    - Forms
        - Choice Based Conjoint(CBC) Analysis: Most Common, Customer chooses most preferred full profile product amongst 3-4 options.
        - Adaptive Conjoint Analysis(ACA): Each customer is asked different set of questions which are dynamically decided based on their responses.

- Full Profile: Full suite of options are presented to the customer and their preference is sought.
- Menu-based: Customer is shown a list of attributes with associate prices, Customer then chooses what they want in their ideal product.

- Applications
  - Marketing: Once Attributes most preferred by customers are known, these can be highlighted in communication channel. Consumers may differ in their choices of preferred attributes and hence CA can help in segmenting the market.
  - Product Development: The product dev team can focus on refining these attributes.
  - Pricing: Organization can decide to price the product based on level of attribute present. May also reveal consumers willingness to pay(WTP) for particular attributes.
- The Process: By defining products as a collection of attributes and having individual consumer react to a number of alternatives. One can infer each attributes importance and most desired level for each customer.
- Conjoint Problem
  Product options are represented as points in multi-attribute space. Different consumer correspond to different "ideal points" that denote their "most preferred". Consumer is supposed to prefer the option which is "closer" to this ideal. As a measure of distance normally either the Euclidean metric or the weighted Euclidean metric is used.
  J = $\{1, 2, 3, ..., n\}$ = set of options
  P = $\{1, 2, ..., t\}$ = n options are described in terms of t dimensions(number of attributes)
  $Y_j = \{y_{jp}\}$ = pre specified location of jth option in the t dimensional space
  $X = \{x_p\}$ = ideal point of consumer
  $d_j^u = \sqrt{\Sigma(y_{jp} - x_p)^2}$ = Unweighted distance, $d_j^w = \sqrt{\Sigma w_p(y_{jp}x_p)^2}$ = Weighted distance
  Weights are non negative for all attributes, $s_j = (d_j^w)^2$ = Squared distance
  $\Omega = \{j, k\}$ denote the set of ordered pairs
  Constraint: $s_k \geq s_j$, B = Poorness of fit = $\Sigma(s_j - s_k)^+ = \max(0, s_j - s_k)$
- Final Formulation
  $a_{jkp} = y_{kp}^2 - y_{jp}^2 \; \forall (j, k) \epsilon \Omega$, $A_p = \Sigma a_{jkp}$
  $b_{jkp} = -2(y_{kp} - y_{jp})$, $D_p = \Sigma b_{jkp}$
  $V = \{v_p\} = \{w_p x_p\}$
  $z_{jk} = \max[0, -[\Sigma w_p a_{jkp} + \Sigma v_p b_{jkp}]]$
  Min Σz_{jk}
  Subject to: $\Sigma w_p a_{jkp} + \Sigma v_p b_{jkp} + z_{jk} \geq 0$
  $\Sigma w_p A_p + \Sigma v_p D_p = 1$
- Statistical Method - Linear Regression
  - Traditionally CA is just a multiple regression problem. Customers ratings for the product concepts form the dependant variables, characteristics of the product are independent variable.
  - Estimated Betas associated with the independent variable are the utilities.
  - R-Square for the regression characterizes the internal consistency of the respondent.
  - Use one hot encoding then remove one column to eliminate col linearity.
  - Part worth: Level utilities for attributes. Total worth of product is calculated from multiple attributes and multiple levels of attributes together, utility values for separate parts pf the product are part worth.
  - Importance of each attribute: Calculate the range of part worth, then find total of the ranges, then the importance = $\frac{range}{total}$