Data Analysis Course

Analysis Design Document (Version-1)

Venkat Reddy

Data Analysis Course

- ## Data analysis design document
  - Introduction to statistical data analysis
  - Descriptive statistics
  - Data exploration, validation & sanitization
  - Probability distributions examples and applications
  - Simple correlation and regression analysis
  - Multiple liner regression analysis
  - Logistic regression analysis
  - Testing of hypothesis
  - Clustering and decision trees
  - Time series analysis and forecasting
  - Credit Risk Model building-1
  - Credit Risk Model building-2

Note

- This presentation is just class notes. The course notes for Data Analysis Training is by written by me, as an aid for myself.

- The best way to treat this is as a high-level summary; the actual session went more in depth and contained other information.

- Most of this material was written as informal notes, not intended for publication

- Please send questions/comments/corrections to venkat@trenwiseanalytics.com or 21.venkat@gmail.com

- Please check my website for latest version of this document

-Venkat Reddy

Contents

- Background, Objective & Scope
- Understanding Data, Data Cleaning & Audit
- Overall summary & Summary by various segments
- Benchmark Analysis, Tracking basic metrics, KPIs
- Control charts , trends & forecasting
- Multivariate analysis & segmentation
- Driver analysis

Data Analysis Course
Venkat Reddy

In scope & Out of scope

- Background
- What is the objective of the project
- What is in scope of the project?
- Are there any data related issues which will make some analysis impossible, hence out of scope?

Data exploration, Data validation & Data sanitization

- Data exploration- Get a feel of the data
- Data validation - Is the data precise?
- Data Sanitization – What if there are some inaccuracies in the data
  - Missing Value Treatment
  - Outlier Treatment Identification & Treatment

Overall summary & Summary by various segments

- Descriptive analysis of objective variable
- Descriptive statistics of other important variables
- Univariate analysis of important fields
- Data visualization of variables
- Analysis across various segments or cuts of the population
- Bivariate analysis & visualizations
  - Analysis with more than two variables
  - Frequencies, means etc., considering combination of variables
  - Correlations and simple regressions

Benchmark Analysis, Tracking derived metrics & KPIs

- Derived variables
- Key processing indicators
  - Ratios & deviations etc.,
- Comparison vs target & average
- RAG- Red Green Amber charts and Dashboards

Control charts & trends, forecasting

- Tracking of important metrics over time
- 1.5 σ control charts
- Time series forecasting of future vales

Multivariate analysis & segmentation

- Finding the groups or segments in the population that are behaving alike
  - Segments with respect to objective
  - Overall segments

10

Details later

Driver analysis

- Regression analysis for finding the most impacting drivers
- Most influencing factors on the objective variable
- Quantifying the impact of each factor & comparison of factors

Details later

11

Venkat Reddy Konasani

Manager at Trendwise Analytics

venkat@TrendwiseAnalytics.com

21.venkat@gmail.com

 +91 9886 768879