Course Syllabus Part I
DSC 530 Data Exploration and Analysis

3 Credit Hours

Course Description
This course introduces complex techniques needed for profiling and exploring data. Students use programming and statistics-based inference to ask and answer insightful questions of data.

Course Prerequisites:
Recommended DSC 510 and DSC 520

Course Objectives

Students who successful complete this course should be able to:

1. Perform exploratory data analysis using programming techniques and statistics-based inference.
2. Analyze datasets to ask and answer insightful questions of data.
3. Evaluate datasets to ensure appropriate quality.
4. Construct a portfolio of data science projects.

Grading Scale

93 – 100% = A	87 – 89% = B+	77 – 79% = C+	67 – 69% = D+
90 – 92% = A-	83 – 86% = B	73 – 76% = C	63 – 66% = D
	80 – 82% = B-	70 – 72% = C-	60 – 62% = D-
			0 – 59% = F

Topic Outline

I. Data Science Process
   a. Importing Data
   b. Exploring Data
   c. Modeling Data
   d. Visualizing Data

II. Python Basics
   a. Installing Tools
   b. Variables
   c. Functions
   d. Packages/Modules

III. Data Quality
   a. How was data collected
   b. What data transformations occurred
   c. When does data need to be fixed
   d. Metadata

IV. Traditional Data Profiling vs Exploratory Data Analysis (EDA)
   a. Data Profiling for Data Warehouses and traditional reporting
      i. Completeness Analysis
      ii. Uniqueness Analysis
      iii. Values Distribution Analysis
      iv. Range Analysis
      v. Pattern Analysis
   b. Exploratory Data Analysis for data science
      i. Statistical Approach
      ii. EDA Process
      iii. Data Mining

V. Exploratory Data Analysis Hands On
   a. Distributions
   b. Probability Mass Functions
   c. Cumulative Distribution Functions
   d. Modeling Distributions
   e. Probability Density Functions
   f. Relationships between Variables
   g. Estimation
   h. Hypothesis Testing
   i. Linear Least Squares
   j. Regression
   k. Time Series Analysis
   l. Survival Analysis
   m. Analytic Methods

VI. Data Modeling Basics
   a. Normalization
   b. Data Cardinality