BELLEVUE UNIVERSITY

Course Syllabus Part I
DSC 540 Data Preparation

3 Credit Hours

Course Description

Much like life, the data humans produce is infinitely variable in its structure, presentation, and scale. This course prepares students for this infinite variety of data. Students use Python, SQL, and other tools to acquire, prepare, clean, and automate dataset creation.

Course Prerequisites

DSC 510 or equivalent and recommend DSC 530

Course Objectives

Students who successful complete this course should be able to:

1. Explain the lifecycle of a data analysis project.
2. Perform techniques to acquire data from various sources.
3. Cleanse, analyze and automate data for various processing needs.
4. Combine multiple data sources together for analysis.
5. Extract meaning from disparate and large datasets.
6. Construct questions that lead to deeper analysis.
7. Present data with purposeful visualizations to tell the story.

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 Flow Framework
- II. Dynamics of Data Wrangling
  - A. Ingesting Data
  - B. Describing Data
  - C. Building an Optimized Dataset
- III. Why Python for Data Wrangling?
- IV. Data Structures
  - A. CSV
  - B. JSON
  - C. XML
  - D. PDF Parsing in Python
- V. Data Profiling
  - A. Syntactic Profiling
  - B. Semantic Profiling
- VI. Acquiring & Storing Data
  - A. Where to find Data
  - B. Storing your Data
  - C. Databases (SQL/NoSQL)
- VII. Data Transformations
  - A. Structuring
  - B. Data Cleanup
  - C. Enriching
- VIII. Data Exploration & Analysis
  - A. Exploring Data
  - B. Analyzing Data
- IX. Data Presentation
  - A. Visualizing Data
  - B. Presentation Tools
- X. Web Scaping
  - A. What to scrape and how
  - B. Reading a webpage with Beautiful Soup
  - C. Screen Scrapers & Spiders
- XI. Data Wrangling: APIs
  - A. Rest vs Streaming
  - B. Rate Limits
  - C. Tiered Data Volumes
- XII. Automation
  - A. Why Automate?
  - B. Steps to Automate
  - C. Tools to Automate
  - D. Simple vs Large Scale Automation
  - E. Monitoring
- XIII. Roles & Responsibilities
- XIV. Data Wrangling Tools
  - A. Excel
  - B. SQL
  - C. Trifacta Wrangler