Seneca | SCHOOL OF SOFTWARE DESIGN & DATA SCIENCE

Course Addendum

Semester**: 2227-Fall 2022** Subject Code: **BDA300** Section**: NAA**
Subject Title: **Data preparation**
Professor: **Rani Gnanaolivu**
E-mail: **rani.gnanaolivu@senecacollege.ca**
Office Hours**:**

Approved by: _____

Kathy Dumanski, Chair, School of Software Design and Data Science

Please read this addendum to the general course outline carefully. It is your guide to the course requirements and activities.
Please refer to the course outline for learning outcomes, course description and text and materials.

Please also visit sdds.senecacollege.ca for key information on courses, graduation requirements, transfer credit, and more from the School of Software Design and Data Science.

Assessment Summary

Assignments	30%
Labs	35%
Tests	20%
Final	15%

Course Policies

- Achieve a grade of 50% or better on the weighted average of the tests and final assessment.
- Grading Policy: http://www.senecacollege.ca/about/policies/grading-policy.html)

Academic Policies:
http://www.senecacollege.ca/about/policies/academics-and-student-services.html

TENTATIVE WEEKLY SCHEDULE
Fall 2021

Week	Topic or Skill	Reading	Assessment	Weight
Week 1 **Sep 5 - 9**	Pandas DataFrame Basics	Ch01: pandas for everyone	**Lab 1: Manipulating Pandas data Frames**	**5%**
Week 2 **Sep 12 - 16**	Pandas data structures	Ch02: Pandas for everyone	**Lab 2: Manipulating Pandas Data structures**	**5%**
Week 3 **Sep 19 - 23**	Importing and exporting data: - Uploading, Streaming, and Sampling Data - Accessing Data in Structured Flat-File Form - Sending Data in unstructured File Form - Managing Data from Relational Databases - Accessing Data from the Web **Quiz 1**	• Ch06: working with real data. Python for Data Science for Dummies, 2nd Edition • Ch06: Data loading, Storage. Python for data analysis : Data wrangling with pandas, NumPy, and IPython	**Lab 3: data integration** **Quiz1**	**Lab 3: 5%** **Quiz1: 3%**
Week 4 **Sep 26 - 30**	Data assembly -Tidy data - Concatenation (adding rows, columns) - Merging multiple datasets	Ch04: Pandas for everyone		
Week 5 **Oct 3 - 7**	Exploratory Data Analysis - Defining descriptive Statistics for Numeric Data - Counting for Categorical Data - Creating Applied Visualization for EDA - Understanding Correlation - Modifying Data Distributions	Ch13 (B1)	**Lab 4: EDA**	**5%**

Week 6
Oct 10 - 14
	Data Munging:
	- Data types
	- Converting types (to string, to numeric)
	- String and text data
	  - Subsetting and slicing strings
	  - String methods
	  - String formatting
		Ch 07, 08:
		Pandas for everyone		

Week 7 **Oct 17 - 21**	**Midterm Exam**			**14%**
Study Week **Oct. 24 to Oct. 28, 2022**				
Week 8 **Oct 31 - Nov4**	Missing Data - Find and count missing data - Cleaning missing data - Calculations with missing data Outlier detection	Ch05: Pandas for everyone	**Lab 5: Working with missing data**	**7%**
Week 9 **Nov 7 - 11**	Apply - Functions - Apply over a series and dataframe - Column-wise operations - Row-wise operations - Vectorized functions - Lambda functions	Ch09: Pandas for everyone	**Lab 6: Apply**	**8%**
Week 10 **Nov 14 - 18**	GroupBy operations: Split-Apply-Combine	Ch10: Pandas for everyone	**Quiz 2** **Assignment:MS1**	**3%** **5%**

Week 11
Nov 21 - 25
	- Sampling
	- String and text data
	  - Subsetting and slicing strings
	  - String methods
	  - String formatting
			Assignment:MS2	**7%**

Week 12 **Nov 28 – Dec 2**	The datetime data type Misc.	Ch11: Pandas for everyone	**Assignment:MS3**	**15%**

Week 13 Dec 5 - 9	Review		**Assignment:MS4**	**3%**
Week 14 Dec 14	**Final Exam**			**15%**

Textbooks:

B1: Python for Data Science for Dummies, 2nd Edition

By: John Paul Mueller, Luca Massaron

Publisher: John Wiley & Sons © 2019

https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5179652260003226

B2: Pandas for Everyone: Python Data Analysis, First Edition

by Daniel Y. Chen

Publisher: Addison-Wesley Professional

Release Date: December 2017

ISBN: 9780134547046

https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5164012370003226

Required software: Anaconda