Text

Description automatically generated

**Kazakh-British Technical University**

**International School of Economics**

Graphical user interface, application

Description automatically generated with medium confidence

**APPROVED**

**Acting Dean of ISE**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Bigabatova M.**

**«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022**

**SYLLABUS**

Academic Year 2022-2023

Fall 2022

**Discipline:** Programming for Data Science 1

**Credits:** 5 ECTS

**Course/Syllabus designer:** Sergei Lytkin, PhD

**Programming for Data Science**

|  |  |
| --- | --- |
| **Instructor’s name** | Sergei Lytkin |
| **Instructor’s e-mail** | smlytkin@gmail.com |
| **Instructor’s telephone** | (optional) |
| **Class Schedule** | Thursday 12.00-13.50  Saturday 14.00-15.50 |
| **Office Hours** | By appointment |
| **Office** | Room 410a |

**1. Course Overview**

In the last decade the demand for programming skills related to managing and visualizing data has grown remarkably. Python, R and SQL feature consistently in the top skills listed in data science and data analyst jobs. Knowing how to write efficient software code to handle and visualise data is an essential skill for any modern data scientist. This course will cover the main principles of computer programming with a focus on data science applications by following the entire pathway from raw data to databases, data wrangling and visualisation, machine learning frameworks up to software development.

Prerequisites: none

**2. Course Objectives and Intended Learning Outcomes**

**Aims and objectives**

* Gain knowledge on the main principles of programming in the Data science context
* Develop ability to handle and visualise data
* Apply computational thinking in various applications domains
* Provide training in state-of-the-art tools, e.g. SQL, Python, R and Git
* Communicate the data analysis results to stakeholders and share work with people in the Data Science industry

**Learning outcomes**

At the end of the course and having completed the essential reading and activities students should be able to:

* Convert raw data to relational databases such as SQL
* Import data to Python and R, apply data manipulation and visualisation
* Program in Python and R
* Develop software using version control via Git

**3.Textbooks and readings**

1. McKinney W. Python for Data Analysis, 2nd edition O’Reilly (2017)
2. Gutagg J.V. Introduction to Computation and Programming using Python, MIT Press, 2nd edition (2017)
3. Wickham H. and Grolemund G. R for Data Science, 1st edition O’Reilly (2017)
4. Wickham H. Advanced R., 1st edition Chapman & Hall (2015)
5. Rammakrishnan R. and Gehrke J. Database Management Systems, 3rd edition, McGraw Hill (2002)

**4. Week by Week Course Outline**

Week 1 -2

**Block 01 Introduction**

Introduction to Data Science

Source-Code Editors and IDEs for R and Python

Installing and Interacting With R

R Markdown and R Notebooks

Installing and Working With Python

Version Control (GIT)

Repository Hosting Services and Collaboration Platforms: GitHub

Week 3-5

**Block 02 Data**

Structured, Semi-Structured, and Unstructured Data

File Formats for Data Exchange

Import/Export of Data-Exchange Files in R

Data Types in R

Data Structures in R

Data Structures in Python

Working With NumPy and Pandas

Data Exchange File Formats With Python

Week 6-8

**Block 03 Introduction to Relational Database Management Systems**

Database Management Systems

Introduction to SQL and SQLite

Basic SQL/SQLite Syntax and Queries

Creating and Manipulating Databases in R Using DBI

Querying Databases in R Using DBI

Querying Databases in R Using dplyr

Creating and Manipulating Databases in Python

Querying Databases in Python

Week 9-11

**Block 04 Programming Concepts**

Computer Programming

Variables, Control Flow Structures and Functions

Exceptions, Error Handling and Debugging in R

Week 12-14

**Block 5 Variables, Mutability and Aliasing in Python and R**

Variables, Mutability and Aliasing in Python and R

Control Flow Structures in Python

Function and Scope in Python

Exceptions and Error Handling in Python

Debugging in Python

Classes and Programming Paradigms

Week 15

Review

**5. Course Requirements and Grades**

Your grade will be based on:

Class Participation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15%

3 Quizzes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .45%

Final Coursework (week 16) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ... 40%

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Total . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .100%

**6. General Rules**

**Academic Integrity**

Any time you put your name on a piece of work for this source you are asserting that it is your own work. If it is not the case, it is plagiarism. You must report, in proper citation format, the source of information used in a paper. Direct quotes must be denoted by quotation marks and page references. Plagiarism is classed as cheating and will be dealt with according to the policies and procedures of KBTU.

Cheating on exams and assignments will not be tolerated and will be dealt with according to the policies and procedures of KBTU.

**Grading Policy**

At the end of the semester a total final score, which is a cumulative measure of your work throughout the semester is calculated. The final rating will be offered according to the scale of assessments adopted in KBTU.

If a student has less than 30 points out of 60 cumulative prior final examination, the student gets a Fail on the course.

If a student gets less than 50% on the final examination (between 10 and 20 points out of 40), the student gets a FX on the course. The student has a second attempt to pass during Re-Sit Examination period.

If a student misses the final examination (with good reason) the student gets a I (Incomplete) on the course. Student will need to provide necessary documents. Final exam will be re-arranged.

If a student is caught cheating on the final examination - the student gets a Fail on the course.

**Attendance policy**

Very regular attendance is expected. Students who are not willing to attend very regularly should drop the course. Important: if you accumulated more than 20% absences throughout the semester, you automatically fail the class