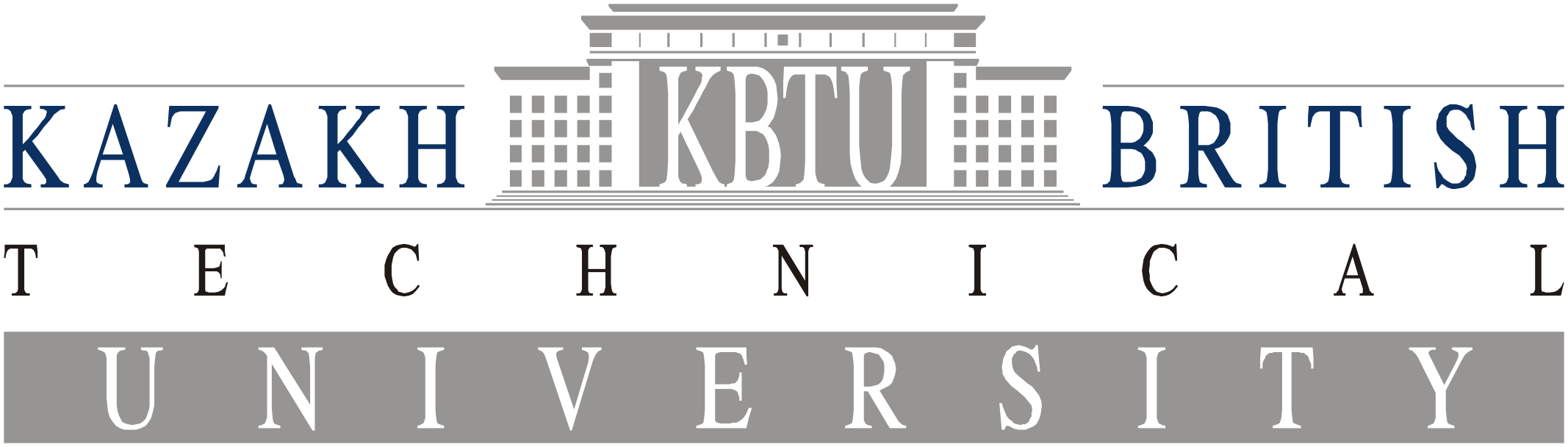
Text

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

Spring 2023

**Discipline:** Programming for Data Science 2

**Credits:** 3

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

|  |  |
| --- | --- |
| **Instructor’s name** | Sergei Lytkin |
| **Instructor’s e-mail** | [s.lytkin@kbtu.kz](mailto:s.lytkin@kbtu.kz) |
| **Instructor’s telephone** | 77054378389 |
| **Class Schedule** | Friday 14.00-17.00 |
| **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
* Get familiar with the most popular machine learning techniques and frameworks
* 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
* Make quantitative predictions via models of machine learning
* 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 06 Introduction to Data Wrangling**

Data Wrangling

Data Wrangling Operations in R

Data Wrangling Operations in Python

Week 3-5

**Block 07 Exploratory Analysis and Data Visualisation**

Basic Graphs and Principles for Visualising Data

Graphics and Data Visualization in R

The Grammar of Graphics and ggplot2

Week 6-8

**Block 08 Graphics and Data Visualisation in Python**

Plotting in Python

Import an Clean Data

**matplotlib** Basics

Line, Bar and Scatter Plots

Histograms, Boxplots and Violin Plots

Heat Map

Exporting Plots

Network Visualizations

Week 9-11

**Block 09 Machine Learning Frameworks**

Machine Learning: Definition and Applications

Elements of a Machine Learning Model

Types of Machine Learning Models

Machine Learning Workflow

Machine Learning Frameworks in R

Categorical Variables in Machine Learning

Machine Learning Frameworks in Python

Week 12-14

**Block 10 Introduction to Software Development**

Software Development Phases

Software Development Life Cycle Methodologies

Developing R packages

Modules and Packages in Python

Documenting Code

Test-Driven Software Development

Week 15

Review

**5. Course Requirements and Grades**

Your grade will be based on:

Class Participation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10%

Quizzes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50%

Final Exam (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