მონაცემთა ანალიტიკა - Python

|  |  |
| --- | --- |
| **კურსის სახელწოდება** | მონაცემთა ანალიტიკა - Python |
| **ინსტრუქტორი** | ნოდარ ოქროშიაშვილი |
| **Email** | nodar.okroshiashvili@btu.edu.ge |
| **კურსის აღწერა** | კურსი მიზნად ისახავს სტუდენტებს შეასწავლოს მონაცემთა ანალიზი პითონის საშუალებით. გააცნოს მონაცემთა ანალიზისთვის საჭირო გავრცელებული მეთოდები და ასწავლოს მათი გამოყენება რეალური ამოცანების გადასაჭრელად. |
| **კურსის შედეგი** | კურსის გავლის შემდეგ სტუდენტს შეეძლება:   * კომპანიის/დამსაქმებლის წინაშე მდგარი პრობლემის იდენტიფიცირება. * საჭირო და შესაბამისი მონაცემების მოძიება და ანალიზისათვის მომზადება. * მონაცემების გაწმენდა და დამუშავება. შედეგების ანალიზი. * მონაცემებზე დაყრდნობით კანონზომიერების შესწავლა და მისი გამოყენება გადაწყვეტილების მისაღებად. * მანქანური სწავლების სხვადასხვა ალგორითმების გამოყენება რეგრესიისა და კლასიფიკაციის ამოცანის გადასაჭრელად. |
| **ლექციების განრიგი** | **TBA** |
| **კურსის წინაპირობა** | * Python I და Python II * სტატისტიკა I |
| **სავალდებულო ლიტერატურა** | **Python for Data Analysis**, Wes McKinney, O'Reilly Media, Inc., 2007  **Python Data Analytics With Pandas, NumPy, and Matplotlib,** Fabio Nelli, Apress Inc, 2018 |
| **დამატებითი ლიტერატურა** | **Python Data Analysis**, Ivan Idris, Packt Publishing, 2014 |

# კურსის შინაარსი

|  |  |  |  |
| --- | --- | --- | --- |
| ლექცია | **მეცადინეობის ფორმა** | **თემა/აქტივობა** | **ლიტერატურა** |
| 1 | ლექცია - 1სთ. | **შესავალი**   * მონაცემთა მეცნიერება * მონაცემთა ანალიტიკა * მანქანური სწავლება * ღრმა სწავლება * მონაცემთა ანალიტიკის/მეცნიერების გამოყენება * რატომ Python მონაცემთა ანალიზისათვის? * სილაბუსის გაცნობა * მნიშვნელოვანი ბიბლიოთეკები * სამუშაო გარემოს გაცნობა | * McKinney, გვ 4-8, 15-46, 51-84 * McKinney, გვ 15-46 * Nelli, გვ 103-121 |
| პრაქტიკული 1სთ-1.5 სთ. | სამუშაო გარემოს მოწყობა და საჭირო ბიბლიოთეკების დაინსტალირება.  მონაცემების წაკითხვა ფაილიდან და ფაილში ჩაწერა.  მონაცემების ზომის და ტიპების შემოწმება, აღწერითი სტატისტიკა, უნიკალური დაკვირვებები, დაკვირვებების დათვლა. | |
| 2 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Introduction to Pandas**   * Pandas Series * Pandas DataFrame * Index, Selection, Filtering * Re-indexing, Dropping Entries, Axis * Sorting and Ranking * Column rename, reorder, insertion, deletion * Arithmetic operations * Statistics with DataFrame | * McKinney, გვ 123-167 * Nelli, გვ 63-95 * [10 minutes to Pandas](https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html) * [Seeing Theory](https://seeing-theory.brown.edu/) |
| 3 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Introduction to NumPy**   * Scalar, Vector, Matrix, NdArray * Indexing and Slicing * Boolean Indexing * Universal Functions * Sorting * Statistical methods * Linear Algebra | * McKinney, გვ 85-122 * Nelli, გვ 35-61 * Idris, გვ 25-58, 59-83 |
| 4 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Data Cleaning and Preparation**   * Finding and filling missing values * Removing duplicate values * Replacing values * Discretization and Binning * Detecting outliers * String manipulations * Variable Transformation | * McKinney, გვ 191-219 |
| 5 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Join, Combine, Reshape DataFrame**   * Combining and Merging * Joining and Concatenation * Reshaping and Pivoting * Groupby * Cross Tabulation * Long to Wide format * Wide to Long format | * McKinney, გვ 221-251, 287-316 * Nelli, 131-165 |
| 6 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Data Visualization**   * Line Plots * Bar Charts * Scatter Plots * Histogram * Pie Chart * Box Plot * The Datazaurus Dozen * Anscombe’s Quartet * Unstructured Quartet | * McKinney, გვ 253-286 * Nelli, გვ 167-235 * [MatPlotLib Tutorials](https://matplotlib.org/tutorials/index.html) * [PyPlot Tutorial](https://matplotlib.org/tutorials/introductory/pyplot.html#sphx-glr-tutorials-introductory-pyplot-py) * [Example Gallery](https://matplotlib.org/gallery/index.html) * [SeaBorn](https://seaborn.pydata.org/tutorial.html) * [Plotly Gallery](https://plotly.com/python/) * [Datazaurus Dozen](https://www.autodeskresearch.com/publications/samestats) * [Anscombe's\_Quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet) |
| 7 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Applied Project 1**   * Fully fledged data analysis * Presenting results and insights * Discussion about improvements * Graphical User Interface for Pandas | * Instructor’s notebook |
| 8 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Time Series**   * Date and Time data types * Dealing with datetime objects * Indexing, Selection, Sub-setting * Periods and Period Arithmetic * Converting Timestamp to period and back * Resampling * Time Shifting * Moving Average | * McKinney, გვ 317-362 |
| 9 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Regression**   * One variable regression * Decision tree regression * Random Forest regression | * Instructor’s notebook * [Ordinary Least Squares](https://www.statsmodels.org/stable/examples/notebooks/generated/ols.html) * [Decision Tree Regressor](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html) * [Random Forest Regressor](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html) |
| 10 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Classification**   * Logistic Regression * Decision Tree classification * Random Forest classification | * Instructor’s notebook * [Logistic Regression](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html) * [Decision Tree Classifier](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html) * [Random Forest Classifier](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html) |
| 11 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Applied Project 2**   * Real world problem to solve * Regression * Classification * Model performance assessment * Room for improvements | * Instructor’s notebook |
| 12 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Association Analysis**   * Introduction to association analysis * Analyzing customer basket * Apriori Algorithm | * Instructor’s notebook * [Market Basket Analysis](https://pbpython.com/market-basket-analysis.html#:~:text=Introduction%20to%20Market%20Basket%20Analysis%20in%20Python%20Posted,called%20association%20analysis%20which%20attempts%20to%20find%20) * [Apriori Algorithm](https://intellipaat.com/blog/data-science-apriori-algorithm/#:~:text=%20How%20Does%20the%20Apriori%20Algorithm%20Work%3F%20,significant%20items%20based%20on%20the%20support...%20More%20) |
| 13 | ლექცია და პრაქტიკული მეცადინეობა - 2 სთ. | **Sentiment Analysis –** If time permit   * Motivation * Summary of Natural Language Processing * Processing textual data * Classifying positive and negative sentiments | * Instructor’s notebook |