



Универзитет „Св. Кирил и Методиј“ во Скопје  
**ФАКУЛТЕТ ЗА ИНФОРМАТИЧКИ НАУКИ И  
КОМПЈУТЕРСКО ИНЖЕНЕРСТВО**

## Програма на обуката за **Бази и анализа на податоци**

Обуката се одвива во времетраење од 6 месеци практична настава и полагање за сертификат во текот на седмиот месец.

Првите два месеци кандидатот се стекнува со општа компјутерска писменост и општи познавања од програмирање. На оваа обука, програмскиот јазик кој што се разгледува за време на почетниот дел од обуката е Python. Кандидатите по завршувањето на основните предмети, можат да изберат во која насока ќе ја продолжат обуката. За обуката за Бази и анализа на податоци има на избор две опции: **Програмирање со бази на податоци** и **Анализа на податоци**.

Во табелите подолу се дадени деталните содржини на предметите кои се опфатени во обуката:

### Општа компјутерска писменост

#### Основи на информатика и компјутерски вештини

Теми	Часови
Вовед во компјутерски вештини и основи на хардвер	4
Оперативни системи - основни операции со практични примери	4
Фајл системи - операции и практични примери	4
Основни програмски пакети и корисни алатки (инсталација и препораки)	4
Користење на програми и корисни алатки (архивирање, операции со слики)	4
Канцелариски алатки (пакети за канцелариско работење)	4
Онлајн комуникациски алатки (емаил, комуникација во реално време)	4
Помош и дијагностика	4

#### Интернет и компјутерска безбедност

Теми	Часови
Основи на Интернет	4
Прегледувачи, HTML и CSS	4
Организација на компјутерски мрежи и TCP/IP протоколно множество	4
Работа со мрежни уреди, основно конфигурирање	4
HTTP, HTTPS и FTP	4
Основи на компјутерска безбедност, заштитни мерки и алатки	4
Сајбер хигиена	4
Заштита на локална мрежа и компјутери од мрежен пристап	4

## Вовед во структурно програмирање

Тема	Часови
Вовед во програмирање, програмски јазици, програмски јазик по избор	2
Начин на запис на броеви во компјутер, концепт на структурно програмирање, променливи, прва програма	2
Операции / аритметички / релациски, логички	4
Контролни структури, акотогаш-инаку, циклуси	4
Алгоритми (правилно алгоритамско размислување и развивање програми)	4
Низи, Матрици	2
Функции 1	2
Стрингови	2
Функции 2 и Рекурзија	4
Датотеки	2
Обработка на интегриран пример	4

## Вовед во Објектно програмирање

Тема	Часови
Објекти и класи (објектно ориентирано моделирање)	8
Дефинирање на објекти, класи, атрибути, методи	4
Објектно ориентиран принцип, наследување, абстракција, полиморфизам, абстрактни класи, преоптоварување на оператори	8
Имплементација на објектно ориентирана апликација (практичен пример)	8
Системи за верзионирање на код и нивно користење	4

## Програмирање со бази на податоци

### Структури и бази на податоци

Теми	Часови
Запознавање со основните структури на податоци (низи, листи), Операции за манипулација со податоците во основните структури на податоци, Запознавање и работа со hash структури, Операции за манипулација со податоците во hash структурите, Користење на hash за дефинирање на речници, Практични примери во Python за користење на овие структури	8
Запознавање и работа со нелинеарни структури (дрва), Операции за манипулација со податоците во дрва, Примери на користење на дрва во Python, Вовед во модели на реалниот свет и нотации за нивно документирање	8

Релациони модели за моделирање на реалниот свет со практични примери, Дефинирање на податочни типови и креирање на табели (DML), Вовед во SQL наредби за со податоци преку практични примери, Дефинирање на надворешни клучеви и референцијален интегритет	8
Работа со повеќе табели, Различни типови на спојувања, Оператори за уни и пресеци од податоци	8

#### Избрани теми од оперативни системи и мрежи

Тема	Часови
Основи на оперативни системи	4
Работа со Command Line Interface околина	4
Основни наредби, работа со датотеки	4
Организација на компјутерски мрежи	4
Мрежни сервиси	4
Хостирање на веб сервиси	4
Работа со сервери и основно конфигурирање	4
Основи на безбедност на софтверските решенија, заштитни мерки и алатки	4

#### Дизајн на бази на податоци и складишта

Теми	Часови
<ul style="list-style-type: none"> <li>· Вовед во модели на реалниот свет и нотации за нивно документирање</li> <li>· EP (УМЛ) дијаграми за моделирање на реалниот свет</li> <li>· Работа на практичен пример – моделирање на база на податоци</li> </ul>	8
<ul style="list-style-type: none"> <li>· Поврзувања на податоците во релациона база на податоци (интегритет на податоците)</li> <li>· Вовед во SQL наредби за внес, промена и бришење на податоци во релациона база на податоци</li> <li>· Работа на практичен пример – креирање на табели и полнење на податоците</li> </ul>	8
<ul style="list-style-type: none"> <li>· Запознавање со SQL стандардот за поставување прашалници до релационите бази на податоци</li> <li>· Спојување на табели</li> <li>· Примена на SQL за генерирање различни типови извештаи со погледи и корелирани прашалници (common table expressions)</li> </ul>	8

<ul style="list-style-type: none"> <li>· Работа на практичен пример – преземање податоци од веќе креирана и пополнета база на податоци</li> </ul>	
<ul style="list-style-type: none"> <li>· Димензионо моделирање – откривање на димензиони табели во постоечки бази на податоци</li> <li>· Дефинирање на факт табели во снегулка и ѕвезда шема преку практични примери</li> <li>· Споредба на оперативни бази на податоци и складишта на податоци и податочни тезги</li> <li>· Денормализација и агрегатни функции</li> </ul>	8

### SQL оптимизација на прашања, складишта и визуелизација

Теми	Часови
Вгнездени прашања и спојувања: споредба преку примери и осврт од аспект на перформанси Променливи Скаларни функции Табеларни функции	8
Аналитички функции за: подредувања, рангирања, кумулативни суми подвижни просеци Групирачки множества Повеќекратни агрегации со Roll Up и Cube	8
Погледи – илустрација преку различни примери Еднократни корелирани прашалници (Common table expressions) Повеќекратни корелирани прашалници (повеќекратни Common table expressions) Споредба од аспект на перформанси и шаблони на употреба на погледни и корелирани прашалници	8
Рекурзивни корелирани прашалници (Recursive common table expressions) Зачувани процедури (Stored procedures)	8
Услови (If-Then-Else logic) Итерации со While loops Итерации со Курсори (Cursors)	8
Процесирање на прашалници (Query processing) Анализа на планови на извршување (Execution plans) на примери од постоечки прашања Индексирање: <ul style="list-style-type: none"> <li>Креирање на примарни индекси</li> <li>Кластерирани индекси</li> <li>Секундарни индекси</li> <li>Дополнителни колони кај секундарни индекси</li> <li>Автогенерирачки колони</li> <li>Globally Unique Identifier колони (GUIDs)</li> </ul> Споредба на планови на извршување	8
Дефинирање на складишта на податоци и тезги на податоци и споредба со оперативни бази на податоци Тек на податоци, архитектури и методологии	8

Димензионални модели Свезда и шема снегулка	
Дизајнирање на модели различни складишта на податоци според примери од оперативни бази на податоци	8
Бавно променливи димензии (slowly changing dimensions): дефинирање на нови мета-колони и дискусија на примената Проширување на моделите на постоечки бази на податоци да поддржуваат бавно променливи димензии Пишување на процедури за екстракција, трансформација и полнење (Extract-Transform-Load, i.e. ETL) на податочните тезги (Data Marts)	8
Концепти за интеграција и миграција на податоци: процеси и техники Правење резервни копии (backups), ресторирање Енкрипција на податоци Безбедносни сертификати	8
Пример со алатки за процесирање (пр. Talend Open Studio, Azure Data Factory, Pentaho Data Integration, итн) Алатки за обогатување и прочистување на податоци (пр. Trifacta Wrangler)	8
Визуализација на податоци Правење dashboards (пр. Power BI, Grafana, Tableau, итн)	8

## Анализа на податоци

### Структури и бази на податоци

Теми	Часови
Запознавање со основните структури на податоци (низи, листи), Операции за манипулација со податоците во основните структури на податоци, Запознавање и работа со hash структури, Операции за манипулација со податоците во hash структурите, Користење на hash за дефинирање на речници, Практични примери во Python за користење на овие структури	8
Запознавање и работа со нелинеарни структури (дрва), Операции за манипулација со податоците во дрва, Примери на користење на дрва во Python, Вовед во модели на реалниот свет и нотации за нивно документирање	8
Релациони модели за моделирање на реалниот свет со практични примери, Дефинирање на податочни типови и креирање на табели (DML), Вовед во SQL наредби за со податоци преку практични примери, Дефинирање на надворешни клучеви и референцијален интегритет	8
Работа со повеќе табели, Различни типови на спојувања, Оператори за унии и пресеци од податоци	8

### Избрани теми од оперативни системи и мрежи

Тема	Часови
Основи на оперативни системи	4
Работа со Command Line Interface околина	4
Основни наредби, работа со датотеки	4
Организација на компјутерски мрежи	4
Мрежни сервиси	4
Хостирање на веб сервиси	4
Работа со сервери и основно конфигурирање	4
Основи на безбедност на софтверските решенија, заштитни мерки и алатки	4

### Бази на податоци за Data Science

Теми	Часови
<ul style="list-style-type: none"> <li>· Вовед во објектно ориентираните концепти на Python (запознавање со базичните структури, листи, множества, dictionary)</li> <li>· Запознавање со Pandas и dataFrames</li> <li>· I/O операции (полнење на dataFrame, конвертирање на основните типови податоци во dataFrame и обратно)</li> <li>· Индексирање и селектирање на податоци</li> </ul>	· 8
<ul style="list-style-type: none"> <li>· Multiindex/advanced index</li> <li>· Слевање, спојување и конкатенирање</li> <li>· Реструктурирање и pivot табели</li> <li>· Групирање и подредување на податоци според една или повеќе колони</li> </ul>	· 8
<ul style="list-style-type: none"> <li>· Запознавање и работа со различни типови податоци</li> <li>· Работа и справување со “missing” податоци</li> <li>· Визуелизација на податоци</li> </ul>	· 8
<ul style="list-style-type: none"> <li>· Поврзување на Python со најчесто користените имплементации на бази на податоци</li> <li>· Објектно-релационо мапирање преку SQLAlchemy и Marshmallow</li> </ul>	· 8

### Анализа на податоци со Python

Теми	Часови
<p>Вовед во курсот: Поставување околина, Jupyter / Colab</p> <ul style="list-style-type: none"> <li>- Што се податоци: Популација, примерок, атрибути (обележје) и објекти</li> <li>- Видови податоци - Основна data wrangling со користење на Python и Pandas,</li> <li>- Примери за студии на случај</li> <li>- Квалитет на податоци</li> </ul>	8
<p>Распределба на честоти на податоци, (табели на фреквенција и хистограми),</p> <ul style="list-style-type: none"> <li>- Библиотеки во Python за работа со податоци (Numpy, scipy)</li> <li>- Емпириска веројатност.</li> <li>- Python за претставување на распределбата на честотите (табели на фреквенција, пити, столбести дијаграми и хистограми)</li> </ul>	8
<p>Дескриптивна анализа:</p> <ul style="list-style-type: none"> <li>- Мерки на централна тенденција и расејување на податоците</li> <li>- Врска помеѓу дескриптивните статистики и графичкиот приказ на податоци, (boxplot, histogram,...)</li> </ul>	8
<ul style="list-style-type: none"> <li>- Студии на случаи за категориски и нумерички податоци</li> <li>- Статистички распределби и Стандардизација на променлива</li> <li>- Изведување заклучоци</li> </ul>	8
<p>Визуелна истражувачка анализа на еднодимензионални податоци со Matplotlib, Seaborn, plotly и Други библиотеки</p>	8
<p>Напредни техники на data wrangling</p> <ul style="list-style-type: none"> <li>- Оценка на податоците визуелно и програмски со помош на Pandas</li> <li>- Проблеми со валкани податоци</li> <li>- Проблемите со квалитетот на податоците и категоризација со метрика: валидност, точност, комплетност, конзистентност и униформност</li> <li>- Чистење податоци со помош на Python и Pandas, Тестирање на код за чистење со помош на Python</li> </ul>	8
<p>Работа со нумерички повеќе-димензионални податоци:</p> <ul style="list-style-type: none"> <li>- Визуелно претставување на односот меѓу две нумерички променливи, scatterplot</li> <li>- Корелација и линеарна регресија</li> <li>- Користење на линеарна регресија за предвидување и кластерирање.</li> </ul>	8
<p>Работа со категориски и нумерички повеќе-димензионални податоци,</p> <ul style="list-style-type: none"> <li>- Користење на кластерирачки графици за приказ на односите помеѓу категориските променливи.</li> <li>- Визуелни графички методи за приказ на односите помеѓу категориски и нумерички променливи.</li> <li>- Логистичка регресија</li> </ul>	8

Студија на случај за визуелизација	8
Повеќе-димензионална регресија <ul style="list-style-type: none"><li>- Визуелизација на повеќедимензионални податоци (користење на бои за претставување на повеќе димензии)</li><li>- Користење на резултатите од повеќе-димензионална линеарна регресија за интерпретирање на коефициентите за неколку предвидувачи.</li></ul>	8
Кластерирање и класификација	8
Работа на практичен и реален пример	8