

Международный институт экономики и финансов
Рабочая программа дисциплины «Python для сбора и анализа данных»
2018-2019 учебный год

Разработчики программы

Щуров И.В., к.ф.-м.н., доцент, ischurov@hse.ru
Тамбовцева А.А., ассистент, atambovtseva@hse.ru

Лекции и семинары ведет

Тамбовцева А.А., ассистент, atambovtseva@hse.ru

1. Цели освоения дисциплины

Целями освоения дисциплины «Python для сбора и анализа данных» являются овладение навыками программирования на языке Python, овладение методами сбора и обработки данных для решения социально-экономических задач.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен учиться, приобретать новые знания, умения, в том числе в области, отличной от профессиональной	УК-1	МЦ	Владеет навыками программирования в Python	Практические задания на семинарах	Домашние задания, проект
Способен собрать и проанализировать исходные данные, необходимые для расчета экономических и социально-экономических показателей, характеризующих деятельность хозяйствующих субъектов	ПК-7	СД	Владеет навыками обработки и визуализации качественных и количественных данных в Python	Практические задания на семинарах, мини-проекты	Домашние задания, проект
Способен осуществлять сбор, анализ и обработку статистических данных, информации, научно-аналитических материалов, необходимых для решения поставленных экономических задач	ПК-11	СД	Находит необходимую информацию в сети Интернет, демонстрирует навыки работы с базами данных	Лекции и семинары	Домашние задания

3. Место дисциплины в структуре образовательной программы

Настоящая дисциплина читается на 3 курсе бакалавриата образовательной программы «Экономика» и является факультативной дисциплиной.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Базовые навыки работы с персональным компьютером
- Знания математики и статистики в рамках обязательных курсов «Теория вероятностей и основы статистики», «Математический анализ»
- Знание английского языка, достаточное для чтения учебной литературы и документации

Основные положения дисциплины могут быть использованы в дальнейшем при изучении дисциплин: «Количественные методы в финансах», «Методы работы с данными для экономистов».

4. Тематический план учебной дисциплины

№	Тема	Всего часов	Лекции	Семинары
1	Установка Python 3 (пакет Anaconda). Знакомство с интерфейсом Jupyter Notebook. Anaconda Command Prompt и установка библиотек. Переменные в Python. Типы данных в Python.	4	2	2
2	Списки, кортежи и множества в Python. Цикл <i>for</i> .	4	2	2
3	Управляющие конструкции в Python. Конструкция <i>if-else</i> . Цикл <i>while</i> .	2	1	1
4	Словари в Python. Работа с json-файлами.	2	1	1
5	Устройство функций в Python. Написание пользовательских функций. Исключения. Поиск ошибок в коде и отладка.	2	1	1
6	Работа с файлами в Python: открытие, изменение, сохранение. Работа с базами данных в Python. Библиотека <i>pandas</i> .	4	2	2
7	Визуализация данных средствами библиотеки <i>matplotlib</i> . Знакомство с библиотекой <i>seaborn</i> .	4	2	2
8	Извлечение данных из веб-страниц. Парсинг html-файлов в Python. Управление браузером из Python. Библиотека <i>selenium</i> .	6	4	2
9	Работа с API. Пример работы с API ВКонтакте и Twitter.	4	2	2
	ИТОГО	32	17	15

5. Формы контроля знаний студентов

Тип контроля	Форма контроля	Параметры
Текущий	Домашнее задание	Письменная работа (набор задач по программированию)
Итоговый	Проект	Проект с последующей устной защитой.

Оценка за **домашнее задание** формируется как округленное среднее арифметическое оценок за текущие домашние задания (способ округления: арифметический).

Проект представляет собой написание корректно работающей программы прикладного назначения с последующей устной защитой в виде презентации. Проект включает в себя следующие части: программа Python (файл с расширением *.py* или файл Jupyter Notebook с расширением *.ipynb*), pdf-файл с описанием назначения и принципов работы программы (документация) и презентация, включающая публичную демонстрацию работы программы. Проект обязательно должен включать использование навыков, полученных в результате прохождения разделов 6-9, то есть выгрузку данных из html-файла, сохранение данных в датафрейм *pandas* с последующей их обработкой, визуализация данных средствами *matplotlib* или других библиотек.

6. Критерии оценки знаний, навыков

Домашние задания представляют собой набор задач по пройденным темам, которые выполняются на компьютере, в Python. Домашние задания и проект оцениваются по 10-ти балльной шкале.

Работа студентов оценивается по следующим критериям: работоспособность и корректность кода программы (код должен запускаться без ошибок и выдавать ожидаемый результат), использование рассмотренных в курсе средств и методов, соответствие требованиям преподавателя (формат и срок сдачи заданий). Задачи, для решения которых предоставлен неработающий код (код, который невозможно запустить из-за наличия грубых ошибок/опечаток), не засчитываются, даже если при этом зафиксирован верный результат.

7. Образовательные технологии

Занятия по курсу включают лекции и семинарские занятия. Лекции и семинары проходят в компьютерном классе.

8. Порядок формирования оценок по дисциплине

Домашние задания

Преподаватель оценивает текущие домашние задания студентов. Домашние задания предполагают решение задач по программированию по пройденной теме. Оценки за домашние задания студента преподаватель выставляет в рабочую ведомость. Общая оценка по 10-ти балльной шкале за самостоятельную работу ($O_{дз}$) определяется как округленное до целого среднее арифметическое оценок, полученных за домашние работы (способ округления: арифметический).

С учетом разного уровня подготовки студентов домашние задания представляют собой набор задач разной сложности, которые весят разное количество баллов. В начале каждого домашнего задания указано общее число баллов, которое можно набрать, чтобы получить оценку 10. Таким образом, студент может выбирать, какие задачи ему решить. Например, если за все домашнее задание нужно набрать 8 баллов, чтобы получить оценку 10, студент может решить 8 задач по одному баллу или 4 задачи по одному баллу и одну задачу на 4 балла. Кроме того, студент может решить больше задач, чтобы потренироваться. Все задачи будут проверены и оценены, но получить оценку выше 10 за домашнее задание невозможно.

Домашние задания, сданные после срока, оцениваются с использованием *понижающих коэффициентов*: опоздание в пределах часа – штраф 10% от полученной оценки, в пределах суток – штраф 20%, в пределах недели – штраф 50%. Домашние задания, сданные через неделю после указанного срока и позже, не принимаются и не оцениваются.

Проект

Итоговый проект выполняется индивидуально или в паре. В исключительных случаях допустимо выполнение проекта в группе из 3 человек. При коллективной защите проекта все члены группы должны быть подготовлены и готовы ответить на вопросы по проекту.

Итоговый проект по курсу оценивается по 10-ти балльной шкале ($O_{\text{проект}}$). Оценка за проект складывается из следующих компонентов (каждый компонент также оценивается в 10-ти балльной шкале):

- Описание проекта, включающее назначение проекта, принцип работы программы, источник используемых данных, планируемые средства визуализации и первичного анализа данных ($O_{\text{план}}$). Сдается не менее, чем за неделю до устной защиты проекта.
- Сама программа (файл *.ру* или *.ipynb*), которая оценивается по следующим критериям: работоспособность кода, соответствие используемых инструментов заявленной цели, применение при ее написании навыков, изученных в разделах 6-9 плана курса (см. пункт 5), корректное оформление кода. ($O_{\text{прог}}$).
- Документация к программе: pdf-файл с описанием функционала программы, требований к данным, подаваемых на вход, источников данных, результата исполнения программы, формата данных на выходе. ($O_{\text{док}}$).
- Презентация программы – устная защита проекта в любом формате, обязательно включающая демонстрацию работы программы. Оценивается по следующим критериям: наличие объяснения цели программы, ее принципа работы и получаемых результатов, способность корректно ответить на вопросы по работе программы в рамках тем, изученных на курсе, объяснить ограничения работы программы и потенциальные проблемы. ($O_{\text{през}}$).

Общая оценка за проект определяется по следующей формуле:

$$O_{\text{проект}} = 0.15 * O_{\text{план}} + 0.4 * O_{\text{прог}} + 0.15 * O_{\text{док}} + 0.3 * O_{\text{през}}.$$

Итоговая оценка за проект округляется (способ округления: арифметический).

Академическая этика

Если при проверке работ (текущий и итоговый контроль) установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется.

9. Накопленная и результирующая оценки

Накопленная оценка по дисциплине рассчитывается по формуле:

$$O_{\text{накопленная}} = O_{\text{дз}},$$

где $O_{\text{дз}}$ — оценка за домашнее задание.

В диплом выставляется результирующая оценка по учебной дисциплине:

$$O_{\text{результатирующая}} = 0.6 * O_{\text{накопленная}} + 0.4 * O_{\text{проект}}$$

Способ округления результирующей оценки по учебной дисциплине: арифметический.

В случае пересдачи накопленная оценка студента сохраняется, а отчетность по курсу производится в том же формате, то есть в виде проекта с последующей защитой.

10. Учебно-методическое и информационное обеспечение дисциплины

Базовый учебник

W. McKinney. Python for Data Analysis. Oreilly Media. 2012. 466 p.

Основная литература

M.Lutz. Learning Python. Oreilly Media. 2013. 1540 p.

Дополнительная литература

Z.Shaw. Learn Python 3 the Hard Way. 2014. 287 p.

Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

- Python 3 (Anaconda)
- Jupyter Notebook

11. Материально-техническое обеспечение дисциплины

Лекции и семинары проводятся в компьютерном классе. Преподавателю для проведения занятий необходим проектор. Студентам во время работы рекомендуется использовать свои ноутбуки. Из программного обеспечения необходимы Python 3 и Jupyter Notebook.