



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины Бизнес аналитика в среде R и Python для
Общеуниверситетского пула МАГОЛЕГО

Правительство Российской Федерации

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет Экономических наук

Программа дисциплины «Бизнес аналитика в среде R и Python»

для Общеуниверситетского пула МАГОЛЕГО

Автор программы:

Поляков К. Л., к.т.н., доцент, polyakov.kl@hse.ru

Москва, 2015

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.



Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов, выбравших из Общеуниверситетского пула МАГОЛЕГО дисциплину «Бизнес аналитика в среде R и Python».

Программа разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ

1 Цели освоения дисциплины

Целями освоения дисциплины освоение методов анализа данных, возникающих в ходе хозяйственной деятельности коммерческих организаций, формирования навыков использования свободно распространяемого программного обеспечения (R и Python) для информационно-технологической поддержки указанного анализа, получение необходимых знаний для разработки аналитических приложений в средах R и Python.

2 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать аналитические потребности коммерческих организаций, вычислительные возможности сред R и Python, а также созданных на их основе приложений и библиотек, основы программирования в указанных средах.
- Уметь формализовать задачу из предметной области, сформулировать требования к данным, подобрать адекватные методы их анализа, выполнить расчеты с использованием сред R и Python или приложений созданных на их основе или создать собственные приложения и проинтерпретировать полученные результаты.
- Иметь навыки самостоятельного анализа данных, возникающих в ходе профессиональной деятельности, а также разработки приложений в средах R и Python.

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по ФГОС/ НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
Способен предлагать концепции, модели, изобретать и апробировать способы и инструменты профессиональной деятельности	СК - М2	Формализация задач из предметной области	Практическая работа на семинарах
Способен анализировать, верифицировать, оценивать полноту информации в ходе профессиональной деятельности, при необходимости восполнять и синтезировать недостающую информацию и работать	СК - М6	Подбор и сбор данных для решения практической задачи. Анализ качества и актуальности данных.	Самостоятельные исследовательские проекты



Компетенция	Код по ФГОС/НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
в условиях неопределенности			
Способен оформлять, представлять и докладывать результаты выполненной работы.	ИК-М3.2 2_2_2.5.2	Подготовка презентаций и отчетов	Выступления на семинарах
Способен порождать принципиально новые идеи и продукты, обладает креативностью, инициативностью.	СЛК-М8	Интерпретация результатов анализа данных	Практическая работа на семинарах

3 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к общеуниверситетскому пулу МАГОЛЕГО.

Изучение данной дисциплины базируется на дисциплинах «Основы теории вероятностей и математической статистики», «Микроэкономика».

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- Методы политического проектирования и сценарирования
- Методы социологических исследований
- Методы симуляционного моделирования
- Политическое прогнозирование
- Методы анализа политических рисков

4 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	Часть 1. Аналитические потребности коммерческой организации. Система стратегических и оперативных аналитических задач.	28	5	5		18
2	Организация работы в вычислительных средах R и Python. Основы программирования. Предварительный анализ данных.	14	7	7		
2	Часть 2. Бизнес-аналитика в среде R.	36	9	9		18
3	Часть 3 . Бизнес-аналитика в среде Python.	36	9	9		18
	Итого:	114	30	30		54



5 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Кафедра	Параметры **
		1	2	3	4		
Текущий (неделя)	Контрольная работа						
	Эссе						
	Реферат						
	Коллоквиум						
	Домашнее задание			*	*		Учебно – научные проекты по изученным темам.
Промежуточный	Зачет						
	Экзамен						
Итоговый	Зачет/ Экзамен [Оставьте нужное]				*		Письменный экзамен 90 мин.

5.1 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на семинарских и практических занятиях на основании активности и результативности участия в работе семинара. Оценки за работу на семинарских и практических занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем - $O_{аудиторная}$.

Преподаватель оценивает самостоятельную работу студентов на основании результатов выполнения учебно-научных исследовательских мини проектов. Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за самостоятельную работу определяется перед промежуточным или итоговым контролем – $O_{сам. работа}$.

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{накопленная,k} = 0.4 * O_{ауд} + 0.6 * O_{домашняя работа} \quad k=1,2$$

$$O_{результ} = 0.7 * (O_{накопл,1} + O_{накопл,2}) + 0.3 * O_{экз}$$

Способ округления накопленной оценки промежуточного (итогового) контроля в форме экзамена: в пользу студента.

ВНИМАНИЕ: оценка за итоговый контроль **блокирующая**, при неудовлетворительной итоговой оценке она равна результирующей.



6 Содержание дисциплины

Часть 1. Аналитические потребности коммерческой организации. Система стратегических и оперативных аналитических задач.

- 1.1. Понятие бизнес-процесса. Сбалансированная система показателей Нортон и Каплана. 5 часов
- 1.2. Аналитические задачи стратегического и оперативного управления. 5 часов

Часть 2. Организация работы в вычислительных средах R и Python. Основы программирования. Предварительный анализ данных.

- 2.1. Организация работы в среде R. 5 часов
- 2.2. Организация работы в среде Python. 5 часов
- 2.3. Типология данных в бизнес-аналитике. Задачи и методы предварительного анализа данных. 4 часа

Часть 3. Бизнес-аналитика в среде R.

- 3.1. Традиционный эконометрический анализ в R. Возможности основных пакетов. 12 часов
- 3.2. Многомерный анализ данных в R. Методы машинного обучения. 12 часов
- 3.3. Анализ временных рядов в R. 12 часов

Часть 4. Бизнес-аналитика в среде Python.

- 4.1. Управление данными с использованием библиотеки Pandas 4 часа
- 4.2. Графическое описание данных с использованием библиотеки Matplotlib 4 часа
- 4.3. Организация вычислений в среде IPython 4 часа
- 4.4. Инструменты машинного обучения scikit-learn 6 часов
- 4.5. Традиционная эконометрика в библиотеке Stats Models 6 часов

7 Учебно-методическое и информационное обеспечение дисциплины

7.1 Базовый учебник

Паклин Н., Орешков В. (2013) Бизнес –аналитика: от данных к знаниям. Питер

7.2 Основная литература

1. Berk R. A. (2008) Statistical Learning from a Regression Perspective. Springer
2. Venables W. N., Ripley B. D. (2002) Modern Applied Statistics with S. Springer
3. Venables W. N., Ripley B. D. (2000) S Programming. Springer
4. Марк Лутц, Программирование на Python (том 1, 2), Москва, Символ-Плюс, 2011
5. Марк Саммерфилд, Программирование на Python 3. Подробное руководство, Москва, Символ-Плюс, 2009

7.3. Дополнительная литература

1. Chambers J.M., Cleveland W. S., Kleiner B., Tukey P. A. (1983) Graphical methods of data analysis. Wadsworth&Brooks/Cole
2. Myatt G. J. (2007) Making Sense of Data. A practical guide to exploratory data analysis and data mining. John Wiley & Sons, Inc.
3. Han J., Kamber M. (2006) Data Mining: Concepts and Techniques, Elsevier Inc.



4. Maddala G. S. (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press

7.3 Программные средства

Среда R

Среда Python