

МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение высшего образования

«Санкт-Петербургский государственный экономический университет»



Анализ данных

Рабочая программа дисциплины

Направление подготовки/ Специальность 38.03.06 Торговое дело

Направленность (профиль) программы/ Специализация Коммерция и электронная торговля

Уровень высшего образования Бакалавриат

Форма обучения очная

Год набора 2023

Составитель(и):
к.физмат.н, Леора Светлана Николаевна

Часов по учебному плану	144	Виды контроля в семестрах: Экзамен: семестр 5
в том числе:		
контактная работа	48	
самостоятельная работа	60	
практическая подготовка	0	
часов на контроль	36	

Распределение часов дисциплины:

Семестр:	5
Вид занятий	Часы
Лекционные занятия	20
Практические занятия	28
Лабораторные работы	
Итого аудиторных часов	48
Самостоятельная работа	60
Часы на контроль	36
Итого академических часов	144
Общая трудоемкость в зачетных единицах	4

Санкт-Петербург
2023

СОДЕРЖАНИЕ

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	3
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	3
3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	3
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*	3
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	4
5.1 Рекомендуемая литература	4
5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства	5
5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД).....	5
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	5
7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	7
8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ	8
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ.....	9
1.1 Контрольные вопросы и задания к промежуточной аттестации	9
1.2 Темы письменных работ.....	10
1.3 Контрольные точки	10
1.4 Другие объекты оценивания	10
1.5 Самостоятельная работа обучающегося	10
1.6 Шкала оценивания результата	10

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель:	Получение навыков использования методов и моделей анализа данных и соответствующего компьютерного инструментария для решения экономических задач.
--------------	---

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина Б1.О Анализ данных относится к обязательной части Блока 1.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Код и наименование компетенции выпускника	Код и наименование индикатора достижения компетенций	Планируемые результаты обучения по дисциплине
ОПК-6 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.	ОПК-6.2 - Использует методы и программные средства для сбора, обработки и анализа данных для решения задач профессиональной деятельности	<p>Знать: принципы работы современных информационных технологий.</p> <p>Уметь: использовать их для решения задач профессиональной деятельности.</p> <p>Владеть: методами и программными средствами для сбора, обработки и анализа данных.</p>
УК-1 - Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2 - Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации	<p>Знать: системный подход для решения поставленных задач.</p> <p>Уметь: осуществлять поиск, критический анализ и синтез информации.</p> <p>Владеть: методами решения проблемной ситуации на основе критического анализа доступных источников информации.</p>

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*

Номер и наименование тем и/или разделов/тем	Содержание дисциплины	Объем дисциплины (академические часы)			
		Контактная работа			СРО
		ЗЛТ	ПЗ	ЛР	
Тема 1. Разведочный анализ данных	Типы и методы представления данных различной природы. Визуализация данных. Анализ основных свойств данных, нахождение в них общих закономерностей, распределений. Одномерный и многомерный анализ данных.	4	5		12

Тема 2. Базовые статистические методы анализа данных.	Описательные статистики. Проверка статистических гипотез. Корреляционный анализ. Анализ таблиц сопряженности.	4	5		12
Тема 3. Обработка данных различной природы.	Работа с пропусками и выбросами. Нормализация и стандартизация данных. Обработка категориальных признаков.	4	6		12
Тема 4. Методы понижения размерности исходных данных	Метод главных компонент. Метод T-SNE. Метод UMAP.	4	6		12
Тема 5. Кластерный анализ.	Кластеры. Метрики близости. Методы объединения кластеров. Иерархический кластерный анализ. Дендрограммы. Метод K-means. Метод DBSCAN. Методы оценки качества кластеризации. Метод силуэта.	4	6		12
Контроль:					36
Всего по дисциплине:		20	28	0	60

*ЗЛТ – занятия лекционного типа, ПЗ – все виды занятий семинарского типа, кроме лабораторных работ, ЛР – лабораторные работы, СРО – самостоятельная работа обучающегося

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Рекомендуемая литература

Библиографическое описание издания (автор, заглавие, вид, место и год издания, кол. стр.)	Электронные ресурсы
Ниворожкина Л.И. Статистические методы анализа данных : учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.] ; под общ. ред. Л.И. Ниворожкиной . — Электрон. дан. - М.; Издательский Центр РИОР: ООО "Научно-издательский центр ИНФРА-М", 2016. – 333 с.	http://znanium.com/go.php?id=556760
Анализ данных : учебник для вузов / В. С. Мхитарян [и др.]; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2021. — 490 с. — (Высшее образование).	URL: https://urait.ru/bcode/469022
Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2020. — 174 с. — (Высшее образование).	https://urait.ru/bcode/450262
Гуриков С.Р. Основы алгоритмизации и программирования на Python : учебное пособие. — Электрон. дан. – М.: ФОРУМ: ИНФРА-М, 2019. – 343 с.	http://znanium.com/go.php?id=970143
Казанский, А. А. Прикладное программирование на Excel 2019 : учебное пособие для среднего профессионального образования / А. А. Казанский. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2021. — 171 с.	https://urait.ru/bcode/470405

5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства

- 7-Zip
- LibreOffice
- ОС Альт образование 10
- Jupyter Notebook
- Python

5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД)

№	Наименование СПБД/ ИСС
1.	Электронная библиотека Grebennikon.ru – www.grebennikon.ru
2.	Научная электронная библиотека eLIBRARY – www.elibrary.ru
3.	Научная электронная библиотека КиберЛеника – www.cyberleninka.ru
4.	База данных ПОЛПРЕД Справочники – www.polpred.com
5.	База данных OECD Books, Papers & Statistics на платформе OECD iLibrary www.oecd-ilibrary.org
6.	Справочная правовая система КонсультантПлюс (инсталлированный ресурс СПБГЭУ или www.consultant.ru)
7.	Справочная правовая система «ГАРАНТ» (инсталлированный ресурс СПБГЭУ или www.garant.ru)
8.	Информационно-справочная система «Кодекс» (инсталлированный ресурс СПБГЭУ или www.kodeks.ru)
9.	Электронная библиотечная система BOOK.ru - www.book.ru
10.	Электронная библиотечная система ЭБС ЮРАЙТ – www.urait.ru
11.	Электронно-библиотечная система ЗНАНИУМ (ZNANIUM) – www.znanium.com
12.	Электронная библиотека СПБГЭУ– opac.unescon.ru

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для реализации данной дисциплины имеются специальные помещения для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ) групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.

Помещения оснащены оборудованием и техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование учебных аудиторий, перечень	Адрес (местоположение) учебных аудиторий
Ауд. 403 Лаборатория "Лабораторный комплекс" Специализированная мебель и оборудование: Учебная мебель на 40 посадочных мест, рабочее место преподавателя, доска меловая - 1 шт., трибуна - 1 шт. Моноблок Acer Aspire Z1811 Intel Core i5-2400S@2.50GHz/4Gb/1Tb - 1 шт., Компьютер I3-8100/8Гб/500Гб/ Philips224E5QSB - 13 шт., Мультимедийный проектор NEC ME401X - 1 шт., Колонки JBL(белые) - 2 шт., Экран с электроприводом Screen Media Champion 203x153cm. MW 4:3. 4-уг. корпус - 1 шт., Микшер-усилитель ТА-1120 - 1 шт., Беспроводная точка доступа/UNI FI AP PRO/Ubiquiti - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	192007, г. Санкт-Петербург, ул. Прилукская, д. 3, лит. А
Ауд. 413 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации), оборудована мультимедийным комплексом. Специализированная мебель и оборудование: Учебная мебель на 28 посадочных мест, рабочее место преподавателя, трибуна - 1 шт., доска аудиторная - 1 шт. Компьютер intel Core i3-8100S/8Gb/1Tb / Philips224E5QSB - 14 шт., Моноблок Acer Aspire Z1811 Intel Core i5-2400S@2.50GHz/4Gb/1Tb - 1 шт., Экран проекцион. Projecta Compact Electrol 153x200 см MATTE White S - 1 шт., Мультимедийный проектор NEC NP-ME402X - 1 шт., Колонки Hi-Fi PRO MASKGT-W- (2шт.) - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	192007, г. Санкт-Петербург, ул. Прилукская, д. 3, лит. А
Ауд. 321 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации), оборудована мультимедийным комплексом. Специализированная мебель и оборудование: Учебная мебель на 32 посадочных мест, рабочее место преподавателя, доска меловая - 1 шт. Переносной мультимедийный комплект: Ноутбук HP 250 G6 1WY58EA, Мультимедийный проектор LG PF1500G. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	192007, г. Санкт-Петербург, ул. Прилукская, д. 3, лит. А

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Приступая к изучению дисциплины, обучающемуся необходимо ознакомиться со следующими документами:

- учебно-методической документацией;
- локальными нормативными актами, регламентирующими основные вопросы организации и осуществления образовательной деятельности, в том числе регламентирующие порядок проведения текущего контроля успеваемости и промежуточной аттестации обучающихся;
- графиком консультаций сотрудников профессорско-преподавательского состава.

Уровень и глубина освоения дисциплины определяются активной и систематической работой обучающихся на лекционных занятиях, занятиях семинарского типа, выполнением самостоятельной работы, в том числе в части выделения наиболее значимых и актуальных проблем для дальнейшего изучения. Особым условием качественного освоения дисциплины является эффективная организация труда, позволяющая распределить учебную нагрузку равномерно в соответствии с графиком учебного процесса.

При подготовке к учебным занятиям обучающимся предоставляется возможность посещения консультаций сотрудников профессорско-преподавательского состава СПбГЭУ согласно расписанию, установленному в графике консультаций.

Аудиторная и внеаудиторная работа обучающихся должна быть направлена на формирование:

- фундаментальных основ мировоззрения обучающихся и естественнонаучного познания;
- базисных знаний, соответствующих направлению подготовки и заявленной профессиональной области, формирующих целевую и профессиональную основу для подготовки кадров;
- профессиональных компетенций ориентированных на удовлетворение потребностей рынка труда;
- индивидуальной траектории посредством освоения уникального набора профессиональных компетенций дополняющих компетентностную модель обучающегося, за счет ориентации на конкретные профессиональные специализированные области знаний, определяемые представителями рынка труда;
- метанавыков обучающихся, таких как: командная работа и лидерство, анализ данных, цифровые навыки, разработка и реализация проектов, межкультурное взаимодействие.

8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

В целях освоения учебной программы дисциплины инвалидами и лицами с ограниченными возможностями здоровья Университет обеспечивает:

- для инвалидов и лиц с ограниченными возможностями здоровья по зрению: размещение в доступных для обучающихся, являющихся слепыми или слабовидящими, местах и в адаптированной форме справочной информации о расписании учебных занятий; присутствие ассистента, оказывающего обучающемуся необходимую помощь; выпуск альтернативных форматов методических материалов (крупный шрифт или аудиофайлы);

- для инвалидов и лиц с ограниченными возможностями здоровья по слуху: надлежащими звуковыми средствами воспроизведение информации;

- для инвалидов и лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата: возможность беспрепятственного доступа обучающихся в учебные помещения, туалетные комнаты и другие помещения кафедры, а также пребывание в указанных помещениях.

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечены печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья. Образование обучающихся с ограниченными возможностями здоровья может быть организовано как совместно с другими обучающимися, так и в отдельных группах или в отдельных организациях.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1.1 Контрольные вопросы и задания к промежуточной аттестации

- 1 Постановка задачи анализа данных. Основные понятия. Типы задач. Примеры прикладных задач.
- 2 Назовите известные вам типы измерительных шкал.
- 3 Оценивание параметров и построение статистических выводов.
- 4 Корреляционный анализ многомерной генеральной совокупности, его назначение и условия применения.
- 5 Применение таблицы сопряженности и критерия хи-квадрат.
- 6 Методы обработки категориальных переменных. Кодирование категориальных переменных.
- 7 Алгоритмы построения решающего дерева. Обработка пропусков в данных.
- 8 Методы поиска локальных аномалий в наборе данных.
- 9 Методы выявления выбросов. Обработка выбросов. Границы Тьюки.
- 10 Методы понижения размерности исходных данных. Основные характеристики.
- 11 Методы t-SNE и UMAP для понижения размерности данных.
- 12 Метод главных компонент. Постановка задачи. Ковариационная матрица исходных признаков.
- 13 Отбор факторов в методе главных компонент. Факторные нагрузки. Оценки качества модели. Интерпретация полученных результатов.
- 14 Основные числовые характеристики и свойства оптимальности главных компонент.
- 15 Условия применимости метода главных компонент. Графический подход.
- 16 Для каких исследовательских целей используется кластерный анализ? Определение кластера.
- 17 Метрики близости в задачах кластеризации. Методы объединения кластеров.
- 18 Методы иерархического кластерного анализа.
- 19 Метод ближайшего соседа. Метод k ближайших соседей. Метод взвешенных ближайших соседей. Выбор весовой функции.
- 20 Методы оценки качества кластеризации.
- 21 Проверка нормальности распределения в пакете Python.
- 22 Математические операции в пакете Python.
- 23 Основные структуры данных в пакете Python: векторы, матрицы, таблицы данных.
- 24 Использование факторов для категориальных переменных в пакете Python.
- 25 Логические операции в пакете Python.
- 26 Создание собственных функций в пакете Python.
- 27 Работа с библиотеками в пакете Python.
- 28 Векторизация, функции apply-семейства в пакете Python.
- 29 Импорт/экспорт в пакете Python.
- 30 Визуализация в пакете Python.
- 31 Найдите все элементы заданного вектора, наиболее близкие к 12. в пакете Python.
- 32 Базовая графика в пакете Python.
- 33 Библиотека statsmodels.
- 34 Напишите функцию, вычисляющую факториал, в пакете Python.
- 35 Найдите в заданном векторе второй элемент, кратный 5. В пакете Python.
- 36 Кластерный анализ в Python.
- 37 Визуализация Matplotlib в Python.
- 38 Визуализация seaborn в Python.
- 39 Визуализация plotly в Python.
- 40 Визуализация в Pandas.

1.2 Темы письменных работ

Рабочей программой дисциплины не предусмотрено.

1.3 Контрольные точки

Номер контрольной точки	Тип контрольной точки	Способ проведения	Номера тем
1	Аналитическая работа	с помощью технических средств и информационных систем	1-3
2	Аналитическая работа	с помощью технических средств и информационных систем	4-5
3	Текущий контроль	с помощью технических средств и информационных систем	1-5

1.4 Другие объекты оценивания

Рабочей программой дисциплины не предусмотрено.

1.5 Самостоятельная работа обучающегося

Наименования самостоятельной работы	Номера тем
Выполнение домашних заданий	1-5
Подготовка к лекционным и практическим занятиям	1-5
Подготовка к экзамену	1-5

1.6 Шкала оценивания результата

Шкалы оценивания и процедуры оценивания результатов обучения **по дисциплине** регламентируются Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по программам высшего образования и Положением о балльно-рейтинговой системе. Для оценки сформированности результатов обучения по дисциплине используется **балльно-рейтинговая система успеваемости обучающихся**:

Формой итогового контроля по дисциплине является экзамен (или дифференцированный зачет), итоговая оценка формируется в соответствии со шкалой, приведенной ниже в таблице:

Баллы	Оценка
<=54	неудовлетворительно
55-69	удовлетворительно
70-84	хорошо
>=85	отлично

Шкала оценивания результата

2 (балл до 54)	Демонстрирует непонимание проблемы. Многие требования, предъявляемые к заданию не выполнены. Демонстрируется первичное восприятие материала. Работа незакончена и /или это плагиат.
3 (балл 55-69)	Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых, к заданию выполнены. Владение элементами заданного материала. В основном выполненный материал понятен и носит целостный характер.
4 (балл 70-84)	Демонстрирует значительное понимание проблемы обозначенной дисциплиной. Все требования, предъявляемые к заданию выполнены. Содержание выполненных заданий раскрыто и рассмотрено с разных точек зрения.
5 (балл 85-100)	Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию выполнены. Продemonстрировано уверенное владение материалом дисциплины. Выполненные задания носят целостных характер, выполнены в полном объеме, структурированы, представлены различные точки зрения, продемонстрирован творческий подход.