

Документ подписан простой электронной подписью Информация о владельце: ФИО: Таскаев Сергей Валерьевич Должность: Ректор Дата подписания: 18.03.2025 14:53:16 Уникальный программный идентификатор: 04c19ed8bfb98f3b6cb77a486b9a8788b8722727	МИНИСТЕРСТВО НАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)	Рабочая программа дисциплины "Сбор, анализ и предобработка данных" по направлению подготовки (специальности) 01.03.02 "Прикладная математика и информатика" направленности (профилю) Прикладная математика и искусственный интеллект ФГБОУ ВО «ЧелГУ»	стр. 1
---	---	---	--------

## **Рабочая программа дисциплины (модуля)\***

**Сбор, анализ и предобработка данных**

Направление подготовки (специальность)

01.03.02 Прикладная математика и информатика

Направленность (профиль)

Прикладная математика и искусственный интеллект

Присваиваемая квалификация (степень)

бакалавр

Форма обучения

очная

Год(ы) набора

2024

\*Рабочая программа дисциплины (модуля) адаптирована для инклюзивного обучения инвалидов и лиц с ограниченными возможностями здоровья

Челябинск 2023 г.



## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре ОПОП
3. Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)
4. Объем дисциплины (модуля)
5. Структура и содержание дисциплины (модуля)
6. Фонд оценочных средств
  - 6.1. Перечень видов оценочных средств
  - 6.2. Типовые контрольные задания и иные материалы для текущей аттестации
  - 6.3. Типовые контрольные вопросы и задания для промежуточной аттестации
  - 6.4. Критерии оценивания
7. Учебно-методическое и информационное обеспечение дисциплины (модуля)
  - 7.1. Рекомендуемая литература
  - 7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"
  - 7.3. Перечень информационных технологий
8. Материально-техническое обеспечение дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Специальные условия освоения дисциплины обучающимися с инвалидностью и ограниченными возможностями здоровья



### 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели:

узнать основные приемы сбора, предобработки, анализа набора данных, а также устройство систем информационного поиска.

Задачами дисциплины являются изучение современных технологий поисковых систем, методов анализа и предварительной обработки данных, а также приобретение навыков программной реализации изучаемых методов с использованием различных программных сред.

Краткое содержание дисциплины:

В курсе рассматриваются основные понятия информационного поиска, технологии поисковых систем, этапы подготовки данных для машинного обучения, методы анализа и предварительной обработки временных рядов.

### 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Цикл (раздел) ОПОП: Б1.В.1.02

#### 2.1 Требования к предварительной подготовке обучающегося:

Нет

#### 2.2 Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

Анализ требований и проектирование систем искусственного интеллекта

Администрирование и проектирование хранилищ больших данных

Математическая статистика

Высокопроизводительные параллельные вычисления на кластерных системах

Практика по получению профессиональных умений и опыта профессиональной деятельности

Современные архитектуры глубоких искусственных нейронных сетей

Введение в искусственный интеллект и нейронные сети

### 3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

**ПК-3: Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта**

**Уметь:**

осуществлять сбор исходной информации с использованием платформ данных (облачных и внутрикорпоративных).

**Владеть:**

практическим опытом сбора информации с использованием платформ данных (облачных и внутрикорпоративных).

#### **ПК-10: Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта**

**Знать:**

методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных;

**Уметь:**

выявлять и исключать из массива данных ошибочные данные и выбросы;  
осуществлять разметку структурированных и неструктурированных данных.

**В результате освоения дисциплины обучающийся должен**

**3.1 Знать:**

3.1.1 методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных.

**3.2 Уметь:**



Рабочая программа дисциплины "Сбор, анализ и предобработка данных" по направлению подготовки (специальности) 01.03.02 "Прикладная математика и информатика" направленности (профилю) Прикладная математика и искусственный интеллект ФГБОУ ВО «ЧелГУ»

стр. 4

3.2.1 осуществлять сбор исходной информации с использованием платформ данных (облачных и внутрикорпоративных), выявлять и исключать из массива данных ошибочные данные и выбросы;

3.2.2 осуществлять разметку структурированных и неструктурированных данных.

### 3.3 Владеть:

3.3.1 сбора информации с использованием платформ данных (облачных и внутрикорпоративных).

## 4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ)

<b>Общая трудоемкость</b>	<b>4 ЗЕТ</b>
Часов по учебному плану : 144 в том числе : контактная работа: 72,5 в том числе : аудиторные занятия : 64 ИКР: 8,5 самостоятельная работа : 71,5	Виды контроля в семестрах:  зачеты с оценкой 3

## 5. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Литература
<b>Раздел 1. Сбор данных. Информационный поиск.</b>				
1.1	Основные понятия информационного поиска. Поисковые системы. Понятие датасета для машинного обучения. /Лек/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
1.2	Лабораторная работа 1. Сбор данных, восстановление пропусков, замена отсутствующих данных, удаление дубликатов, удаление объектов с отсутствующими данными. /Лаб/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 2. Предварительная обработка данных. Технологии анализа данных.</b>				
2.1	Python для обработки и анализа данных. Установка и настройка Jupyter Notebook, переменные и типы данных, строки, списки, цикл for, вложенные списки, условный оператор, цикл while, функции, словари, Pandas для анализа данных. /Лек/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
2.2	Предварительный анализ данных. Описательная статистика. Восстановление пропусков в данных. Методология CRISP-DM для подготовки данных. /Лек/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
2.3	Лабораторная работа 2. Использование Excel для описательной статистики. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
2.4	Лабораторная работа 3. Использование Python для описательной статистики. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 3. Корреляционный анализ данных.</b>				
3.1	Корреляционный анализ. Основные понятия. /Лек/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
3.2	Корреляционный анализ качественных и количественных связей. /Лек/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1



3.3	Канонические корреляции. /Лек/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
3.4	Лабораторная работа 4. Корреляционный анализ качественных признаков. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
3.5	Лабораторная работа 5. Корреляционный анализ количественных признаков. /Лаб/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
3.6	Контрольная работа 1. Теоретический материал по тематикам: классификация данных, пространственные, временные данные, корреляционный анализ. /Лаб/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 4. Регрессионный анализ данных.</b>				
4.1	Регрессионный анализ. Основные понятия. /Лек/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
4.2	Виды регрессионных моделей. /Лек/	3	6	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
4.3	Лабораторная работа 6. Линейная регрессионная модель. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
4.4	Лабораторная работа 7. Нелинейная регрессионная модель. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 5. Анализ временных данных.</b>				
5.1	Анализ временных данных. Основные понятия. Методы сглаживания временных данных и моделирования тенденции развития. /Лек/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
5.2	Лабораторная работа 8. Анализ временных данных. /Лаб/	3	4	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
5.3	Контрольная работа 2. Теоретический материал по тематикам: регрессионные модели, временные ряды. /Лаб/	3	2	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 6. Самостоятельная работа студента</b>				
6.1	Подготовка к лабораторным работам №1-8 /Ср/	3	41,5	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
6.2	Подготовка к дифференцированному зачету /Ср/	3	30	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1
<b>Раздел 7. Иная контактная работа</b>				
7.1	Консультации и промежуточная аттестация /ИКР/	3	8,5	Л1.1 Л1.2 Л1.3Л2.1Л3.1 Э1

## 6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

### 6.1. Перечень видов оценочных средств

(Текущий контроль) Лабораторная работа №1.  
(Текущий контроль) Лабораторная работа №2.  
(Текущий контроль) Лабораторная работа №3.  
(Текущий контроль) Лабораторная работа №4.  
(Текущий контроль) Лабораторная работа №5.  
(Текущий контроль) Лабораторная работа №6.



(Текущий контроль) Лабораторная работа №7.  
(Текущий контроль) Лабораторная работа №8.  
(Текущий контроль) Контрольная работа №1.  
(Текущий контроль) Контрольная работа №2.  
(Промежуточная аттестация) Дифференцированный зачёт.

### 6.2. Типовые контрольные задания и иные материалы для текущей аттестации

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

### 6.3. Типовые контрольные вопросы и задания для промежуточной аттестации

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

### 6.4. Критерии оценивания

КМ №№ 1-8 (лабораторные работы):

3 балла: Студент отвечает на все теоретические вопросы по теме лабораторной работы. Все задания лабораторной работы выполнены.

2 балла: Студент отвечает с затруднениями на теоретические вопросы по теме лабораторной работы. Все задания лабораторной работы выполнены.

1 балл: Студент отвечает с затруднениями на теоретические вопросы по теме лабораторной работы. Задания лабораторной работы полностью не выполнены.

0 баллов: Студент не отвечает на теоретические вопросы по теме лабораторной работы. Задания лабораторной работы полностью не выполнены.

КМ №№ 9-10 (контрольные работы):

4 балла: Студент ответил на все теоретические вопросы контрольной работы.

3 балла: Студент ответил на 3 вопроса контрольной работы.

2 балла: Студент ответил на 2 вопроса контрольной работы.

1 балл: Студент ответил на 1 вопрос контрольной работы.

0 баллов: Студент не ответил на вопросы контрольной работы.

КМ №11 (дифференцированный зачёт):

Процедура проведения:

Рейтинг обучающегося по дисциплине формируется по результатам текущего контроля. Контрольное мероприятие дифференцированного зачета проводится в очной форме и не является обязательным, однако студент может прийти на дифференцированный зачет и повысить свой рейтинг. Студенту на дифференцированном зачете выдаётся билет. Дается 90 минут для подготовки к ответу. Проводится собеседование по выданным вопросам.

4 балла получает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные билетом и свободно отвечающий на дополнительные вопросы.

3 балла заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в билете задания, но отвечающий на дополнительные вопросы с затруднениями.

2 балла получает студент, допустивший погрешности в ответе на дифференцированном зачете и при выполнении заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.

1 балл ставится студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных билетом заданий.

0 баллов ставится студенту, который не смог выполнить ни одно задание в билете.

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

### 7.1. Рекомендуемая литература

#### 7.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л1.1	Геофили Т.	Глубокое обучение для поисковых систем ( <a href="https://e.lanbook.com/book/140574">https://e.lanbook.com/book/140574</a> )	Москва : ДМК Пресс, 2020	ЭБС
Л1.2	Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н.	Большие данные. Big Data: учебник для вузов ( <a href="https://e.lanbook.com/book/322664">https://e.lanbook.com/book/322664</a> )	Санкт-Петербург : Лань, 2023	ЭБС



	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л1.3	Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., Сиротин В. П.	Анализ данных: учебник для вузов ( <a href="https://urait.ru/bcode/511020">https://urait.ru/bcode/511020</a> )	Москва : Юрайт, 2023	ЭБС

#### 7.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л2.1	Маккинни У.	Python и анализ данных ( <a href="https://e.lanbook.com/book/131721">https://e.lanbook.com/book/131721</a> )	Москва : ДМК Пресс, 2020	ЭБС

#### 7.1.3. Методические разработки

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л3.1	Жуковский О. И.	Информационные технологии и анализ данных: учебное пособие ( <a href="https://e.lanbook.com/book/110351">https://e.lanbook.com/book/110351</a> )	Москва : ТУСУР, 2014	ЭБС

#### 7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Электронно-библиотечная система издательства Лань <a href="https://e.lanbook.com/book/">https://e.lanbook.com/book/</a>
Э2	Электронно-библиотечная система Znanium.com <a href="https://znanium.com/catalog/product/">https://znanium.com/catalog/product/</a>
Э3	Образовательная платформа Юрайт <a href="https://urait.ru/bcode/">https://urait.ru/bcode/</a>

#### 7.3 Перечень информационных технологий

##### 7.3.1 Программное обеспечение

OpenOffice

Python

##### 7.3.2 Профессиональные базы данных и информационно-справочные системы

1. Электронный каталог научной библиотеки ЧелГУ [Электронный ресурс] : база данных / Челяб. гос. ун-т. – Челябинск, 1992.
2. eLIBRARY.RU : научная электронная библиотека : сайт. – Москва, 2000 – . – URL: <a href="https://elibrary.ru">https://elibrary.ru</a> – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.
3. Mathematical Reviews (MR) : реферативная база данных / American Mathematical Society. – URL: <a href="http://www.ams.org/mathscinet/">http://www.ams.org/mathscinet/</a> – Яз. рус., англ. – Режим доступа: для зарегистрир. пользователей ЧелГУ. – Текст : электронный.

#### 8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для реализации дисциплины используются учебные аудитории для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы. Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения. Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий таких как презентации лекций, размещенные в системе Moodle. Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с подключением к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

#### 9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания к СРС.

из них: учебно-методическое обеспечение самостоятельной работы студента:

Методические указания к СРС

#### 10. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ ОБУЧАЮЩИМИСЯ С ИНВАЛИДНОСТЬЮ И ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Освоение дисциплины инвалидами и лицами с ограниченными возможностями здоровья осуществляется с использованием специальных технических средств и голо информационных технологий, предоставляемых Ресурсным учебно-методическим центром по обучению инвалидов и лиц с ограниченными возможностями здоровья ЧелГУ по запросу обучающегося.

1. Мобильные специальные технические средства для лиц с нарушениями зрения: портативный компьютер с вводом/выводом шрифтом Брайля с синтезатором речи «EIBraile-W14J G2»; ноутбуки с программной экранного доступа NVDA; электронные увеличители для удаленного просмотра; видеувеличители портативные; тифлоплеер; цифровые диктофоны.

2. Мобильные специальные технические средства для лиц с нарушениями слуха: система свободного звукового поля со встроенной совместимостью с FM-устройствами; радиоклассы «Сонет-PCM» с передатчиком, заушным индуктором и индукционной петлей; система информационная для слабослышащих переносная «Исток» А2 со встроенным плеером – звуковым информатором; документ-камера; программируемые слуховые аппараты индивидуального пользования.



3. Ассистивные информационные технологии: программное обеспечение экранного доступа с синтезом речи NVDA; программы экранного увеличения; программы речевого синтеза для компьютеров и ноутбуков; программы речевого синтеза для мобильных устройств; экранная клавиатура; экранная лупа.

При необходимости для обучающихся с нарушениями зрения на рабочих местах для проведения практических или лабораторных занятий устанавливается специальное программное обеспечение (программа речевой навигации NVDA, речевые синтезаторы, экранные лупы).

В учебные аудитории обеспечивается беспрепятственный доступ для обучающихся инвалидов и обучающихся с ограниченными возможностями здоровья. В каждой аудитории, где обучаются инвалиды и лица с ограниченными возможностями здоровья, предусматривается соответствующее количество мест для обучающихся с учетом нарушений их здоровья.

Для освоения дисциплины инвалидам и лицам с ограниченными возможностями здоровья предоставляется доступ к печатным источникам, имеющимся в научной библиотеке ЧелГУ, с помощью специальных технических средств; доступ к электронным источникам, представленным в форме электронного документа в фонде научной библиотеки ЧелГУ или электронно-библиотечных системах, с помощью специальных технических и программных средств (рабочее место для незрячего пользователя с программным обеспечением экранного доступа с синтезом речи NVDA, рабочее место с компьютерным роллером и клавиатурой Clevy с большими кнопками и с разделяющей клавиши накладкой).

Учебно-методические материалы для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме шрифтом Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Для инвалидов и лиц с ограниченными возможностями здоровья освоение дисциплины может быть частично или полностью осуществлено с использованием дистанционных образовательных технологий (Moodle, Adobe Connect Pro и пр.).

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья используется индивидуальная работа. Под индивидуальной работой подразумевается две формы взаимодействия с преподавателем: индивидуальная учебная работа (консультации), т.е. дополнительное разъяснение учебного материала и углубленное изучение материала с теми обучающимися, которые в этом заинтересованы, и индивидуальная воспитательная работа. Индивидуальные консультации направлены на индивидуализацию обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или обучающимся с ограниченными возможностями здоровья.

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине обеспечивается выполнение следующих дополнительных требований в зависимости от индивидуальных особенностей, обучающихся:

- а) инструкция по порядку проведения процедуры оценивания предоставляется в доступной форме (устно, в письменной форме, в письменной форме шрифтом Брайля, устно с использованием услуг сурдопереводчика);
- б) доступная форма предоставления заданий оценочных средств (в печатной форме, в печатной форме увеличенным шрифтом, в печатной форме шрифтом Брайля, в форме электронного документа, задания зачитываются ассистентом, задания предоставляются с использованием сурдоперевода);
- в) доступная форма предоставления ответов на задания (письменно на бумаге, набор ответов на компьютере, письменно шрифтом Брайля, с использованием услуг ассистента, устно).

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями. Эти средства могут быть предоставлены ЧелГУ или могут использоваться собственные технические средства. При необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на задания, процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Проведение процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья допускается с использованием дистанционных образовательных технологий.

## Контрольное мероприятие 1 – Лабораторная работа 1

Сбор данных, восстановление пропусков, замена отсутствующих данных, удаление дубликатов, удаление объектов с отсутствующими данными.

1. Используя платформы данных, найдите дата сет, который имеет
  - пропуски,
  - дубликаты,
  - ошибочные данные.
2. Проведите процедуру очистки данных.

## Контрольное мероприятие 2 – Лабораторная работа 2

### Описательная статистика.

1. Загрузите датасет с Kaggle <https://www.kaggle.com/datasets>  
Минимальные требования к датасету: минимум 3 столбца и 150 строк данных (можно воспользоваться приложенным уже скачанным датасетом).
2. Выясните, сколько у Вас строк, столбцов, пропущенных значений.
3. Задайте как минимум 4 вопроса к своему датасету (пример, сколько человек имеет опыт менее 3 лет, но получает выше средней зар.платы) и постройте разнообразные графики в Excel, визуализирующие ответы на эти вопросы.

## Контрольное мероприятие 3 – Лабораторная работа 3

### Описательная статистика в Python.

1. Воспользуйтесь датасетом из Лабораторной работы 2.
2. Выясните, сколько у Вас строк, столбцов, пропущенных значений с помощью Python.
3. Задайте как минимум 4 вопроса к своему датасету (пример, сколько человек имеет опыт менее 3 лет, но получает выше средней зар.платы) и постройте разнообразные графики, визуализирующие ответы на эти вопросы.
4. С помощью функции `profile_report` из модуля `pandas_profiling` получите основные сведения о Вашем датасете.
5. Визуализируйте с помощью `missingno` матрицу пропусков.

## Контрольное мероприятие 4 – Лабораторная работа 4

### Корреляционный анализ качественных признаков

1. При ранжировании оценок на вступительных экзаменах и средних баллов за экзаменационную сессию получена таблица. Проверьте наличие статистически значимой взаимосвязи между оценками на вступительных экзаменах и средними баллами за экзаменационную сессию по формулам и проверьте с помощью средств Python.

Студент	Оценка на вступительных экзаменах	Средний балл за экзаменационную сессию
А	9	3
Б	6	4
В	3	3
Г	4	2
Д	8	3
Е	3	4
Ж	6	5
З	5	4
И	8	5
К	4	3
Л	5	4
М	4	3
Н	5	4
О	10	5
П	2	4
Р	10	4

## Контрольное мероприятие 5 – Лабораторная работа 5

### Корреляционный анализ количественных данных

1. На основании данных о деятельности шести организаций оценим тесноту связи между прибылью ( $y$ ) и затратами на 1 руб. произведенной продукции ( $x$ ) по формулам и проверьте с помощью средств Python, используя данные таблицы.

	x	y
1	111	0.57
2	114	0.47
3	112	0.5
4	124	0.34
5	94	1.07
6	100	1

## Контрольное мероприятие 6 – Лабораторная работа 6

### Линейная регрессионная модель

1. В таблице представлены данные о себестоимости продукции ( $y$ ) и производительности труда ( $x$ ). Проведите регрессионный анализ зависимости  $y$  от  $x$  в предположении, что имеет место линейная модели вида  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$M\varepsilon_i = 0, M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j \end{cases}$$

$x$	1	2	3	4	5
$y$	6	5	4	4	3
$\hat{y}_i$	5.8	5.1	4.4	3.7	3.0
$y_i - \hat{y}_i$	0.2	-0.1	-0.4	0.3	0

## Контрольное мероприятие 7 – Лабораторная работа 7

### Нелинейная регрессионная модель

1. Постройте и сравните следующие степенные уравнения регрессии
  - а)  $\tilde{y} = \beta_0(x_2)^{\beta_2}$  – функция спроса
  - б)  $\tilde{y} = \beta_0(x_1)^{\beta_1}$  – функция потребления
  - в)  $\tilde{y} = \beta_0(x_1)^{\beta_1}(x_2)^{\beta_2}$  – функция спроса и потребления
  - г)  $\tilde{y} = \beta_0(x_2)^{\beta_2}(x_3)^{\beta_3}(x_4)^{\beta_4}$  – функция спроса с учетом цены на товарозаменители.

$y$  – объем потребления пирожных

$x_1$  – среднедушевой доход

$x_2$  – стоимость 1 кг пирожных

$x_3$  – стоимость 1 кг мороженого

$x_4$  – стоимость 1 кг конфет

Год	$y$	$x_1$	$x_2$	$x_3$	$x_4$
2004	30,8	459,7	39	55,3	79,2
2005	31,2	492,9	37,3	54,7	77,4
2006	33,3	528,6	38,1	63,7	80,2
2007	35,6	560,3	39,3	69,8	80,4
2008	36,4	624,6	37,8	65,9	83,9
2009	36,7	666,4	38,4	64,5	85,5
2010	38,4	717,8	40	70	93,7
2011	40,4	768,2	38,6	73,2	106,1
2012	40,3	843,3	39,8	67,8	104,8
2013	41,8	911,6	39,7	79,1	114
2014	40,4	931,1	52,1	95,4	124,1
2015	41,8	1021,5	48,9	94,2	127,6
2016	40,4	1165,9	58,3	123,5	142,9
2017	40,7	1349,6	57,9	129,9	143,6
2018	42,7	1449,4	56,5	117,6	139,2
2019	44,1	1575,5	63,7	130,9	165,5
2020	46,7	1759,1	61,6	129,8	203,3
2021	50,6	1994,2	58,9	128	219,6
2022	50,1	2258,1	66,4	141	221,6
2023	52,9	2478,7	70,4	168,2	232,6

## Контрольное мероприятие 8 – Лабораторная работа 8

### Анализ временных данных

По данным из таблицы определить коэффициенты показательной трендовой модели  $\tilde{y} = ab^t$

Год	Прибыль
1	73.5
2	80.3
3	82.5
4	87.4
5	90.7
6	98.5
7	102.8
8	114.6
9	120.7

## Контрольное мероприятие 9 – Контрольная работа 1

1. Назовите основные критерии классификации данных.
2. Назовите основные типы переменных в зависимости от шкалы измерения
3. Какая зависимость между переменными называется корреляционной?
4. Как проверить значимость парных и частных коэффициентов корреляции?

## Контрольное мероприятие 10 – Контрольная работа 2

1. В чем смысл метода наименьших квадратов и для чего он используется?
2. Что характеризуют элементы ковариационной матрицы?
3. Каково назначения теста Бокса – Льюнга?
4. Что понимается под автокорреляционной функцией?

Контрольное мероприятие 11 – Дифференцированный зачет

Федеральное государственное автономное образовательное учреждение  
высшего образования «Южно-Уральский государственный университет  
(национальный исследовательский университет)»

Институт естественный и точных наук

Кафедра прикладной математики и программирования

01.03.02 Прикладная математика и информатика

Дисциплина: Сбор, анализ и преобразование данных

1. При ранжировании оценок на вступительных экзаменах и средних баллов за экзаменационную сессию получена таблица. Проверьте наличие статистически значимой взаимосвязи между оценками на вступительных экзаменах и средними баллами за экзаменационную сессию

Студент	Оценка на вступительных экзаменах	Средний балл за экзаменационную сессию
А	9	3
Б	6	4
В	3	3
Г	4	2
Д	8	3

2. По данным из таблицы определить коэффициенты показательной трендовой модели  $\tilde{y} = ab^t$

Год	Прибыль
1	84
2	91
3	100
4	82
5	88
6	86
7	90
8	78
9	81

3. Параметр адаптации в процедуре экспоненциального сглаживания.
4. Дайте содержательную интерпретацию коэффициента регрессии в простейшем линейном уравнении регрессии.

