**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА»**

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

|  |
| --- |
| «УТВЕРЖДАЮ»Декан экономического факультета МГУпрофессор \_\_\_\_\_\_\_\_\_\_\_\_\_ А.А.Аузан«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_2023 год |
|  |

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Наименование дисциплины:**

**Анализ данных в бизнесе: продвинутый уровень**

**Уровень высшего образования:**

МАГИСТРАТУРА

**Направление подготовки:**

38.04.02. МЕНЕДЖМЕНТ

**Форма обучения:**

ОЧНАЯ

Рабочая программа рассмотрена и одобрена

*Учебно-методической комиссией экономического факультета*

(протокол №\_\_\_\_\_\_\_\_\_\_, дата)

Москва 2023

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки магистратуры 38.04.02. Менеджмент

ОС МГУ утвержден решением Ученого совета МГУ имени М.В.Ломоносова от 28 декабря 2020 года, протокол №7

Год (годы) приема на обучение: 2022 и последующие

1. **Место и статус дисциплины в структуре основной профессиональной образовательной программы подготовки магистра**

Статус дисциплины: *вариативная*

Триместр: 3

1. **Входные требования (реквизиты) для освоения дисциплины**

*Для успешного освоения данного курса требуются знания и умения, полученные в следующих дисциплинах:*

* Методы исследования в менеджменте
1. **Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников**

|  |  |  |
| --- | --- | --- |
| **Компетенции выпускников (коды)** | **Индикаторы (показатели) достижения компетенций** | **Планируемые результаты обучения по дисциплине (модулю), сопряженные с компетенциями** |
| ПК-7.Способен оценить потребности организации в информации, готовить аналитические материалы для экспертизы управленческих процессов | ПК-7.И-1. Проводит анализ потребности организации в информации | ПК-7.И-1.З-1 Знает цели и задачи информационного менеджмента, причины возникновения информационных потребностей, методы изучения информационных потребностей |
| ПК-7.И-1.У-1. Умеет анализировать информационные потребности и определять запросы организации в информационно-аналитическом обеспечении  |
| ПК-7.И-1.У-2. Умеет готовить материалы, раскрывающие потребности организации в информации для экспертизы управленческих процессов |
| ПК-11. Способен самостоятельно создавать модели и алгоритмы типовых и нестандартных организационно-управленческих задач по оптимизации внутренней среды и адаптации к требованиям внешней среды для деятельности конкретного функционального подразделения и организации в целом  | ПК-11.И-1. Проводит анализ факторов внешней и внутренней среды деятельности конретного функционального подразделения и организации в целом | ПК-11.И-1.У-1. Умеет собирать и обрабатывать необходимую информацию для проведения анализа внешней и внутренней среды деятельности конкретного функционального подразделения и организации в целом  |
|  | ПК-11.И-2. Разрабатывает модели и алгоритмы типовых и нестандартных организационно-управленческих решений | ПК-11.И-2.У-1. Умеет использовать основные методы анализа и моделирования управленческих решений в условиях неопределённости и недостаточности информации  |
| МПК-3. Способен разрабатывать проектные решения, бизнес-планы и бизнес-модели, применяя технологии анализа данных, а также предложения и мероприятия по реализации разработанных проектов и программ | МПК-3.И-2. Использует современные методы анализа данных в процессе разработки, анализа и реализации проектных решений, бизнес-планов и бизнес-моделей  | МПК-3.И-2.У-1. Умеет применять основные методы анализа данных в процессе разработки и реализации проектных решений, бизнес-планов и бизнес-моделей |
| МПК-4. Способен оценивать результаты реализации стратегий и проектов, проводить аналитические исследования и осуществлять консультационную деятельность в сфере актуальных проблем управления развитием бизнеса на основе анализа данных | МПК-4.И-2. Проводит аналитические исследования в сфере актуальных проблем управления развитием бизнеса | МПК-4.И-2.У-2. Умеет проводить сбор информации о бизнес-проблемах или бизнес-возможностях для целей исследования |

1. **Объем дисциплины по видам занятий**

Объем дисциплины составляет \_3\_ зачетные единицы: \_108\_ академических часов, в том числе \_52\_ академических часов составляет контактная работа с преподавателем, \_56\_ академических часов составляет самостоятельная работа магистранта.

1. **Формат обучения \_** очное, c использованием обучающей среды On.Econ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**
2. **Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий**

|  |  |  |
| --- | --- | --- |
| **Наименование и краткое содержание разделов и тем дисциплины (модуля),** **Форма промежуточной аттестации по дисциплине (модулю)** | **Всего****(часы**) | В том числе |
| **Контактная работа (работа во взаимодействии с преподавателем)** ***Виды контактной работы, часы*** | **Самостоятельная работа обучающегося** ***Виды самостоятельной работы, часы*** |
| Занятия семинарского типа | Групповые консультации | Индивидуальные консультации | **Всего** | Подготовка к аудиторным занятиям | Самостоятельная внеаудиторная работа | **Всего** |
| Тема 1. Загрузка и преобразование, первичное описание данных. | 16 | 4 | - | 4 | **8** |  | 8 | **8** |
| Тема 2. Визуализация данных посредством ggplot2 и plotly  | 32 | 8 | - | 8 | **16** | 2 | 8 | **10** |
| Тема 3. Представление данных в табличном выражении | 16 | 4 | - | 4 | **8** |  | 8 | **8** |
| Тема 4. Принятие решений на основании регрессионных моделей | 16 | 4 | - | 4 | **8** |  | 8 | **8** |
| Тема 5. Принятие решений на основании логистических моделей | 16 | 4 | - | 4 | **8** |  | 8 | **8** |
| Промежуточная аттестация: итоговая индивидуальная работа | 12 | 4 | - | - | **4** |  | 14 | **14** |
| **Итого**  | **108** | **52** | **56** |

**Краткое содержание тем дисциплины**

**Тема 1. Загрузка и преобразование данных**

Знакомство с функционалом RStudio: синтаксис, особенности компиляции, загрузка и импорт данных. Типы данных в R: character, factor, integer, logical, numeric, matrix, data.frame. Операции с разными типами данных. Работа с пропущенными наблюдениями. Описательные статистики (summary, glimpse): минимум, медиана, среднее значение, максимальное значение. Работа с таблицами посредством пакета dplyr (функции filter, select, transmute, mutate, arrange, group\_by, summarise). Объединение таблиц: merge, join (left, right), cbind, rbind. Группировка данных: создание срезов, построение сводных таблиц.

**Основная литература:**

1. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с. Глава 1 – Знакомство с R; Глава 2 – Создание набора данных; Глава 5 – Более сложные способы управления данными.

**Тема 2. Визуализация данных (ggplot2 и plotly)**

Изучение инструментов визуализации данных по ggplot2 и plotly: гистограммы, графики плотности распределения, ящики с усами, скрипичные графики, столбчатые диаграммы, диаграммы частот, тепловая карта, диаграмма рассеивания. Дифференциация графиков в зависимости от типа данных, корректность их использования. Персонализация графика: параметры facets, цвета, единицы измерения.

**Основная литература:**

1. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с. Глава 3 – Начало работы с диаграммами; Глава 6 – Базовые диаграммы; Глава 11 – Диаграммы средней сложности, Глава 16 – Продвинутые графические методы.

**Дополнительная литература:**

1. Материалы по графикам Plotly // URL: <https://plotly.com/r/>
2. R Graph Gallery // URL: <https://r-graph-gallery.com/index.html>
3. Chris Chapman, Elea McDonnell Feit. R for Marketing Research and Analytics. Springer

**Тема 3. Представление данных в табличном выражении**

Использование пакета gt (Grammar of Tables), gt\_summary, tbl\_summary для представления таблиц. Внедрение в статистические тесты, отражение статистических тестов в виде таблицы. Использование спарклайнов (gt\_plt\_sparkline). Изменение группировки. Формирование подитогов в строках.

**Основная литература:**

 1. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с

**Дополнительная литература:**

1. Creating gt tables // URL: <https://gt.rstudio.com/articles/intro-creating-gt-tables.html>
2. Документация по пакету gt // URL: <https://cran.r-project.org/web/packages/gt/gt.pdf>

**Тема 4. Принятие решений в бизнесе на основе регрессионных моделей**

Методы и инструменты проверки гипотез. Построение регрессионных моделей. Проверка на значимость коэффициента и уравнения в целом. Работа с выбросами. Визуализация и интерпретация результатов. Проблемы регрессии: мультиколлинеарность, гетероскедастичность, эндогенность. Выгрузка моделей (использование функции stargazer)

**Основная литература:**

1. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с. Глава 8 – Регрессия, Глава 7 – Основные методы статистической обработки данных.

**Дополнительная литература:**

1. Лагутин М.Б. - Наглядная математическая статистика: учебное пособие - Издательство "Лаборатория знаний" (ранее "БИНОМ. Лаборатория знаний") – 2019
2. Картаев Ф. Введение в эконометрику: учебник. – М.: Экономический факультет МГУ имени М.В. Ломоносова, 2019. – 472 с.
3. Лебедев А.В., Фадеева Л.Н. Теория вероятностей и математическая статистика (4-е издание). — Москва, 2018.

**Тема 5. Принятие решений в бизнесе на основе логистических моделей**

Модели бинарного выбора и их ценность. Принципы работы с бинарными данными. Интерпретация параметров модели. Матрица сопряженности. Прогноз на основе логистических моделей.

**Основная литература:**

1. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с. Глава 13 – Обобщенные линейные модели.
2. **Фонд оценочных средств для оценивания результатов обучения по дисциплине**
	1. **Примеры оценочных средств:**

|  |  |
| --- | --- |
| **Результаты обучения по дисциплине**  | **Виды оценочных средств** |
| МПК-3.И-1.З-1. Знает основные методы разработки проектных решений и бизнес-планов, характеристики и особенности реализации основных бизнес-моделей | Экзамен |
| ПК-11.И-1.У-1. Умеет собирать и обрабатывать необходимую информацию для проведения анализа внешней и внутренней среды деятельности конкретного функционального подразделения и организации в целом | Текущая аттестация-1: выполнение домашнего задания №1Текущая аттестация-2: подготовка результатов по части 1 и части 2 проекта |
| ПК-11.И-2.У-1. Умеет использовать основные методы анализа и моделирования управленческих решений в условиях неопределённости и недостаточности информации  |
| ПК-11.И-2.У-2. Умеет разрабатывать и обосновывать эффективные организационно-управленческие решения с учётом требований внешней и внутренней среды  |
| МПК-2.И-2.У-1. Умеет оценивать экономические и социальные условия осуществления предпринимательской деятельности, выявлять новые рыночные возможности и формировать новые бизнес-модели |
| МПК-3.И-1.У-1. Умеет разрабатывать различные проектные решения, бизнес-планы и бизнес-модели |
| МПК-3.И-2.У-1. Умеет применять основные методы анализа данных в процессе разработки и реализации проектных решений, бизнес-планов и бизнес-моделей |
| МПК-4.И-1.У-2. Умеет критически оценивать результаты реализации стратегий и проектов при помощи количественных и качественных методов анализа данных |
| МПК-4.И-2.У-3. Уметь применять современные методы проведения аналитических исследований и использовать способы сбора, обработки и анализа данных |
| МПК-4.И-3.У-1. Умеет профессионально проводить анализ и обосновать решения по развитию бизнеса по заданиям заказчика |

* 1. **Критерии оценивания (баллы) по дисциплине:**

|  |  |
| --- | --- |
| **Виды оценочных средств** | **Баллы** |
| Домашняя работа №1 (описательные статистики, визуализация по ggplot2, plotly) | 50 |
| Проект – первая часть работы (по темам 1-3) | 25 |
| Проект – вторая часть работы (по темам 4-6) | 50 |
| Проект – защита проекта с презентацией по итогам работы (экзамен) | 25 |

*При наличии блокирующих элементов до промежуточной аттестации указывается порядок проведения пересдач по каждому из таких элементов.*

* 1. **Оценка по дисциплине выставляется, исходя из следующих критериев:**

|  |  |  |
| --- | --- | --- |
| **Оценка** | **Минимальное количество баллов** | **Максимальное количество баллов** |
| *Отлично*  | 127,5 | 150,0 |
| *Хорошо*  | 97,5 | 127,4 |
| *Удовлетворительно*  | 60,0 | 97,4 |
| *Неудовлетворительно*  | 0,0 | 59,9 |

**Примечание:** в случае, если магистрант за триместр набирает менее 20% баллов от максимального количества по дисциплине, то уже на промежуточном контроле (и далее на пересдачах) действует следующее правило сдачи: «магистрант может получить только оценку «Удовлетворительно», и только если получит за промежуточный контроль, включающий весь материал дисциплины, не менее, чем 85% от баллов за промежуточный контроль».

* 1. **Типовые задания и иные материалы, необходимые для оценки результатов обучения:**

На семинарах студенты выполняют **задания**, направленные на закрепление полученных знаний в ходе занятия. Промежуточное задание оценивается в 50 баллов, итоговое задание складывается за счет 3 этапов в разбивке 25/50/25.

**Пример задания для закрепления материала темы 1**

В ходе семинара мы построили сводную таблицу, в которой показано распределение посетителей фитнес-клубов по дням недели. При помощи изученных инструментов библиотеки ggplot2 постройте тепловую карту и плотность распределения, подобрав релевантный критерий bins. Какие выводы можно сделать о загруженности фитнес-клубов?

**Пример домашнего задания**

Совсем скоро — 27 марта 2022 года — пройдёт 94-я церемония награждения премии «Оскар». Имена номинантов на премию уже известны, а до объявления победителей ждать почти месяц. Проанализируйте всех победителей в номинации «лучший фильм» за всю историю премии. Информация о номинантах и победителях «Оскара» за 1928-2020 годы представлена в таблице oscars\_df.csv

**Инструкция к домашнему заданию**

***1. Откройте таблицу и изучите общую информацию о данных***

Скачайте с портала on.econ.msu.ru таблицу oscars\_df.csv. Помните, что для корректной работы файл с данными и файл с проектом должны находиться в одной папке. Изучите общую информацию о таблице и сделайте выводы

***2. Подготовьте данные для анализа***

Определите размерность используемой таблицы данных.

Вспомните основные правила оформления заголовков столбцов. Все ли столбцы в oscars\_df.csv соответствуют требованиям? Приведите их к верному виду.

В таблице есть пропущенные значения. Опишите, какие пропущенные значения вы обнаружили. Почему могли возникнуть эти пропуски? Можно ли их заполнить или лучше удалить данные с пропусками? Объясните, почему вы принимаете то или иное решение.

Соответствует ли содержание всех столбцов их типам? Проведите необходимые замены типов данных, поясняя свои действия.

Сколько дубликатов содержит таблица? Удалите строки-дубликаты. Есть ли скрытые дубликаты в таблице? Например, что в столбце Award есть одни и те же значения, но записанные по-разному: где-то опечатались в количестве букв n, а где-то написали статус фильма со строчной буквы вместо заглавной. Есть ли аналогичные артефакты в других ячейках?

В выводах отметьте не только, сколько именно дубликатов было удалено, но и возможные причины их появления

***3. Ответьте на следующие вопросы:***

*1. Выиграли ли Оскар самые лучшие фильмы по версии рейтинга IMDB?*

Проанализируйте топ-25 фильмов с самым высоким и топ-25 фильмов с самым низким рейтингом IMDB и сделайте выводы, нарисовав соответствующие графики и построив сводные таблицы.

*2. Влияет ли продолжительность фильма на его успех у критиков киноакадемии?*

Сравните (а) среднюю длительность всех фильмов, номинированных на Оскар, (б) среднюю длительность непобедивших лент и (в) среднюю длительность победителей.

Для выполнения п.2 рекомендуем создать отдельные таблицы с выигравшими и проигравшими фильмами

*3. Сколько часов смотрели фильмы, взявшие награду за лучшую киноленту, критики IMDB?*

Создайте новый столбец, который покажет, сколько часов потратили все пользователи IMDB, которые поставили свою оценку за фильм. Какой фильм посмотрели больше всего?

Интерпретируйте все ответы на вышестоящие вопросы: поясните, о чём именно говорит полученный вами результат.

***4. Сделайте общий вывод по проекту***

Сделайте общий вывод про проведённому исследованию: что удалось сделать и к каким заключения прийти?

Что не хватило вам в исследовании? На какие вопросы вы хотели бы получить ответы в рамках анализа предложенных данных (но не можете этого сделать, т.к. вам пока не знакомы соответствующие инструменты анализа)?

**Промежуточная аттестация** проводится в формате **итоговой проверочной работы**. Работа представляет собой применение изученных в ходе семестра инструментов на примере мини-проекта. В ходе написания работы студенты могут пользоваться материалами занятий. В первую очередь оценивается умение студента ориентироваться в инструментах анализа данных и навыки интерпретаций полученных результатов.

* 1. **Методические рекомендации и требования к выполнению заданий:**

**Критерии оценки домашнего задания и итоговой работы**

**1. Правильность и рациональность применения инструментов**
Используете ли вы все изученные в ходе семинаров инструменты работы с данными? Правильно ли вы используете эти методы? Какие методы замены типов данных, обработки пропусков и дубликатов применяете?

**2. Умение анализировать полученные данные**
Как вы описываете найденные в данных проблемы? Соблюдаете ли структуру проекта и поддерживаете аккуратность кода? Какие выводы делаете? Оставляете ли комментарии к шагам?

**3. Оформление работы**
Правильно оформите работу: помимо кода (в ячейках code) добавляйте пояснения к своим действиям (через #). Не забывайте использовать форматирование и заголовки.

1. **Ресурсное обеспечение**
	1. **Перечень основной и дополнительной литературы**

**Основная литература:**

Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.

**Дополнительная литература:**

1. Creating gt tables // URL: <https://gt.rstudio.com/articles/intro-creating-gt-tables.html>
2. Документация по пакету gt // URL: <https://cran.r-project.org/web/packages/gt/gt.pdf>
3. Лагутин М.Б. - Наглядная математическая статистика: учебное пособие - Издательство "Лаборатория знаний" (ранее "БИНОМ. Лаборатория знаний") – 2019
4. Картаев Ф. Введение в эконометрику: учебник. – М.: Экономический факультет МГУ имени М.В. Ломоносова, 2019. – 472 с.
5. Лебедев А.В., Фадеева Л.Н. Теория вероятностей и математическая статистика (4-е издание). — Москва, 2018.
6. Материалы по графикам Plotly // URL: <https://plotly.com/r/>
7. R Graph Gallery // URL: <https://r-graph-gallery.com/index.html>
8. Chris Chapman, Elea McDonnell Feit. R for Marketing Research and Analytics. Springer

**8.2. Перечень лицензионного программного обеспечения**

RStudio

**8.3. Перечень профессиональных баз данных и информационных справочных систем**

Ресурсы институциональной подписки экономического факультета МГУ. Базы данных из открытых источников: статистические службы, базы данных международных организаций и так далее (см. пп. 8.4)

**8.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости)**

1. Google Тренды (<https://www.google.com/trends/explore>)
2. Платформа Kaggle (<https://www.kaggle.com/>)
3. Google Dataset Search (<https://datasetsearch.research.google.com/>)
4. Федеральная служба государственной статистики (<https://www.gks.ru/>)

**8.5. Описание материально-технической базы**

* + - компьютерный класс с проектором и компьютером для проведения занятий;
		- портал экономического факультета МГУ(www.on.econ.msu.ru);
		- доступ студента в институциональную подписку факультета.
1. **Язык преподавания:** Русский
2. **Преподаватель (преподаватели):** Мирзоян Ашот Гамлетович, Щукина Полина Олеговна
3. **Разработчики программы:** Мирзоян Ашот Гамлетович