**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ**

**ЗАДАНИЕ**

для самостоятельной работы по курсу

**«Проектный семинар “Python в науке о данных”»**

Студент(ы) группы БИВ225:

* Якушев Тимофей Павлович, tpyakushev@edu.hse.ru, +7 961 192 3813
* Татаринова Полина Юрьевна, pyutatarinova@edu.hse.ru, +7 910 352 2609
* Пыжов Илья Игоревич, iipyzhov@edu.hse.ru, +7 905 317 3073

Тема работы:

«Приложение для анализа данных кредитных историй заемщиков»

Руководитель:

Полякова Марина Васильевна

Москва 2023

Оглавление

[1. Введение 3](#_Toc136224080)

[1.1. Наименование программы 3](#_Toc136224081)

[1.2. Краткая характеристика области применения 3](#_Toc136224082)

[2. Основания для разработки 3](#_Toc136224083)

[3. назначение разработки 3](#_Toc136224084)

[3.1. Функциональное назначение 3](#_Toc136224085)

[3.2 Эксплуатационное назначение 3](#_Toc136224086)

[4. Требования к программе или программному изделию 3](#_Toc136224087)

[4.1. Требования к функциональным характеристикам 3](#_Toc136224088)

[4.1.1. Требования к составу выполняемых функций 3](#_Toc136224089)

[4.1.2. Требования к организации входных и выходных данных 4](#_Toc136224090)

[4.2. Требование к надежности 4](#_Toc136224091)

[4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы 4](#_Toc136224092)

[4.2.2 Время восстановления после отказа 4](#_Toc136224093)

[4.2.3 Отказы из-за некорректных действий оператора 4](#_Toc136224094)

[4.3 Условия эксплуатации 4](#_Toc136224095)

[4.3.1 Климатические условия эксплуатации 4](#_Toc136224096)

[4.3.2 Требования к видам обслуживания 4](#_Toc136224097)

[4.3.3 Требования к численности и квалификации персонала 4](#_Toc136224098)

[4.4 Требования к составу и параметрам технических средств 4](#_Toc136224099)

[4.5 Требования к информационной и программной совместимости 4](#_Toc136224100)

[4.5.1 Требования к исходным кодам и языкам программирования 4](#_Toc136224101)

[4.5.2 Требования к программным средствам, используемым программам 5](#_Toc136224102)

[4.6 Требование к маркировке и упаковке 5](#_Toc136224103)

[4.7 Требования к транспортированию и хранению 5](#_Toc136224104)

[5. Требования к программной документации 6](#_Toc136224105)

[5.1 Состав программной документации 6](#_Toc136224106)

[5.2 Специальные требования к программной документации 6](#_Toc136224107)

[6. Технико-экономические показатели 6](#_Toc136224108)

[7. Стадии и этапы разработки 6](#_Toc136224109)

[8. Порядок контроля и приёмки 6](#_Toc136224110)

# 1. Введение

## 1.1. Наименование программы

«Приложение для анализа данных потенциальных заемщиков»

## 1.2. Краткая характеристика области применения

Данное приложение разработано для анализа данных (кредитная история, социальный статус, уровень образования и др.) потенциальных заемщиков с целью оценки банком рисков займа и оптимизации процесса решения о выдаче кредита.

# 2. Основания для разработки

Приложение является итоговой работой курса проектного семинара "Python в науке о данных" департамента компьютерной инженерии Московского государственного института электроники и математики (МИЭМ) имени А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики».

В процессе разработки приложения студенты (исполнители) практикуются в использовании языка Python 3.10, его стандартных библиотек, а также NumPy, Pandas, Matplotlib.

# 3. назначение разработки

## 3.1. Функциональное назначение

Приложение предоставляет пользователю (предположительно, банку либо любой другой организации, занимающейся займами) возможность загрузки базы данных о потребителях банковских услуг, построения по ним графиков, выявления зависимостей, а также предсказания возможных значений у новых данных, что фактически означает анализ рисков выдаваемого кредита.

## 3.2 Эксплуатационное назначение

«Приложение для анализа данных потенциальных заемщиков» - приложение, разработанное специально для банков для оптимизации процесса решения о выдачи кредита. Используя данные о потенциальном заемщике, банк с помощью данного приложения сможет оценить безопасность займа, способность заемщика к периодическим выплатам, выявить оптимальные сумму и срок кредита.

# 4. Требования к программе или программному изделию

## 4.1. Требования к функциональным характеристикам

### 4.1.1. Требования к составу выполняемых функций

1. Приложение запускается из командной строки: python <имя главного модуля>.py. При запуске и использовании приложения не должно возникать ошибок – не должно выводиться на консоль сообщений о некорректной работе приложения.
2. Приложение содержит следующие функции:

* Взаимодействие с базой данных (редактирование - добавление, удаление, обновление, а также сохранение внесенных изменений)
* Отчеты:
  + Построение графиков и диаграмм (кластеризованная столбчатая гистограмма по качественной переменной при разных условиях, категоризированная гистограмма распределения количественной переменной при разных условиях, категоризированная диаграмма Бокса-Вискера по количественной переменной и качественному атрибуту и категоризированная диаграмма рассеивания на основе двух количественных переменных и качественного атрибута);
  + Создание текстового отчета на основе базы данных (выведение статистики и подтаблиц, построенных по заданным пользователем характеристикам).

### 4.1.2. Требования к организации входных и выходных данных

Основные входные данные хранятся в файле формата .cvs. и .xls/xlsx  
База данных имеет следующие поля:  
CLIENTNUM, Attrition\_Flag, Customer\_Age, Gender, Dependent\_count, Education\_Level, Marital\_Status, Income\_Category, Card\_Category, Months\_on\_book, Total\_Relationship\_Count, Months\_Inactive\_12\_mon, Contacts\_Count\_12\_mon, Credit\_Limit, Total\_Revolving\_Bal, Avg\_Open\_To\_Buy, Total\_Amt\_Chng\_Q4\_Q1, Total\_Trans\_Amt, Total\_Trans\_Ct, Total\_Ct\_Chng\_Q4\_Q1, Avg\_Utilization\_Ratio

## 4.2. Требование к надежности

### 4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы

У пользователя должны быть установлены все библиотеки, согласно конфигурационному файлу, для корректной реализации работы программы. Приложение должно иметь постоянный, бесперебойный доступ к базе данных. На устройстве пользователя, на котором запускается программа, должно использоваться лицензионное ПО.

### 4.2.2 Время восстановления после отказа

Время на восстановление после отказа, вызванного неисправностью технических средств или сбоем операционной системы, не должно превышать времени, необходимого на устранение неисправностей технических средств и переустановки программных средств.

### 4.2.3 Отказы из-за некорректных действий оператора

После запуска приложения отказ приложения вследствие некорректных действий оператора должен быть исключен.

## 4.3 Условия эксплуатации

### 4.3.1 Климатические условия эксплуатации

Требований к климатическим условиям эксплуатации не предъявляется.

### 4.3.2 Требования к видам обслуживания

Обслуживание не требуется.

### 4.3.3 Требования к численности и квалификации персонала

Для управления системой достаточно одного человека, способного запустить программу.  
Требуемая квалификация – оператор ЭВМ.

## 4.4 Требования к составу и параметрам технических средств

Вся реализация программы осуществляется непосредственно на устройстве пользователя. В состав технических средств должен входить персональный компьютер.  
Его минимальные характеристики:

* процессор ×86 с тактовой частотой не менее 1 ГГц;
* оперативная память объемом, не менее 1 Гб;
* видеокарта, монитор, мышь, клавиатура;
* OC MS Windows 10, MacOS High Sierra

## 4.5 Требования к информационной и программной совместимости

### 4.5.1 Требования к исходным кодам и языкам программирования

Исходный код программы должен быть написан на языке программирования Python3  
Код разработан в соответствии с требованиями «Python Enhancement Proposals (PEP) 8 -- Style Guide for Python Code» и «PEP 257 -- Docstring Conventions»

### 4.5.2 Требования к программным средствам, используемым программам

Требуются следующие программные средства

1. Компьютер с установленной операционной системой
2. Установленный дистрибутив Anaconda не ниже 2022.05
3. Python 3.10
4. В системе должны быть установлены следующие библиотеки для корректной работы программы:
   1. contourpy 1.0.7
   2. cycler 0.11.0
   3. et-xmlfile 1.1.0
   4. fonttools 4.38.0
   5. kiwisolver 1.4.4
   6. matplotlib 3.5.3
   7. numpy 1.21.6
   8. openpy×l 3.1.2
   9. packaging 23.0
   10. pandas 1.3.5
   11. Pillow 9.4.0
   12. pyparsing 3.0.9
   13. python-dateutil 2.8.2
   14. pytz 2022.7.1
   15. six 1.16.0
   16. tzdata 2023.3

## 4.6 Требование к маркировке и упаковке

Программа поставляется в виде zip-архива, требующий распаковывания и установки сторонних библиотек.  
Информационно–аналитическое приложение размещается в стандартной структуре каталогов:

Work <- основной каталог.

Data — содержит базу данных.

Graphics — содержит копии графических отчетов.

Library — содержит библиотеку стандартных (универсальных) функций, разработанных

бригадой, которые могут использоваться для создания других приложений, например, функции

чтения файлов.

Notes — содержит документацию, в нем размещается Руководства пользователя и

разработчика.

Output — содержит копии текстовых отчетов.

Scripts — содержит специализированный модуль и файл с определением параметров

настройки приложения.

## 4.7 Требования к транспортированию и хранению

Специальные требования к транспортировке не предъявляются.

# 5. Требования к программной документации

## 5.1 Состав программной документации

1. Техническое задание
2. Пояснительная записка
3. Руководство пользователя
4. Руководство разработчика
5. Программа и методика испытаний
6. Текст программы

## 5.2 Специальные требования к программной документации

Документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 и ГОСТами к каждому виду документа. Документация и программа также сдается электронном виде в формате .pdf или docx. в архиве формата .гір.

# 6. Технико-экономические показатели

Использование данного приложения позволит пользователю (банку) сократить время на анализ данных потенциального заемщика и оптимизирует процесс выдачи кредита.

# 7. Стадии и этапы разработки

|  |  |
| --- | --- |
| Дата | Контрольная точка |
| 1.02.2023 | Согласование индивидуального проекта |
| 15.03.2023 | Представление первых графиков |
| 10.05.2023 | Статистические отчеты |
| 25.05.2023 | Финальное техническое задание и графический интерфейс |
| 11.06.2023 | Сдача проекта |

# 8. Порядок контроля и приёмки

Производится проверка корректного выполнения программой заложенных в нее функций, т. е. осуществляется функциональное тестирование программы. Также осуществляется визуальная проверка интерфейса программы. Программа и методика испытаний (ГОСТ 19.301-79)”, в котором указывают:

1. перечень функций программы, выделенных в программе для испытаний, и перечень требований, которым должны соответствовать эти функции
2. перечень необходимой документации и требования к Функциональное тестирование осуществляется в соответствии с документом «Программа методы испытаний и обработки информации»;
3. технические средства и порядок проведения испытаний;

Финальная версия информационно-аналитического приложения, «Руководства пользователя» и «Руководства разработчика» выкладывается в Telegram-группу Директором бригады не позднее, чем за 10 дней до начала сессии.

Финальная версия, а также промежуточные версии, предоставляются преподавателю и ассистентам строго в виде архивов каталогов «work», содержащих все компоненты проектов. Название архива –

<Группа(три символа)>\_<Бригада(два символа)>\_<Версия(три символа)>.zip. Архивы должны иметь постоянный на весь срок обучения пароль, который Директор сообщает преподавателю в личном сообщении в Telegram.