**Задача 1:**

В таблице хранятся сведения о состоянии договоров за каждый месяц.

Пример (1 договор):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **id договора** | **Дата заключения договора** | **Дата отчета** | **Сумма выдач** | **Задолженность в валюте приведения** | **Задолженность 30+** | **Задолженность 90+** |
| 1153864318 | 01.01.2019 | 01.01.2019 | 605575.87 | 605575.87 | 0 | 0 |
| 1153864318 | 01.01.2019 | 01.02.2019 | 605575.87 | 605575.87 | 0 | 0 |
| 1153864318 | 01.01.2019 | 01.03.2019 | 605575.87 | 601234.51 | 0 | 0 |

Полные данные приложены в файле:



Необходимо в excel сделать сводную таблицу для формирования винтажа по задолженности 90+ дней:

Пример винтажа (Рисунок 1):

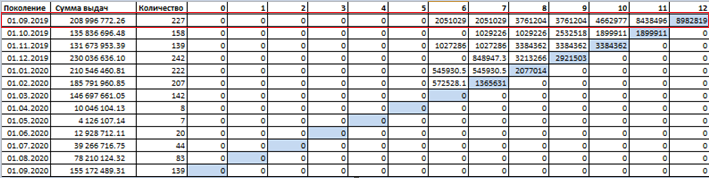


Рисунок 1 – Пример винтажа

Где:

**Поколение** – месяц выдачи кредитов

**Сумма выдач** – сумма выданных кредитов за соответствующий месяц

**Количество** – кол-во выданных договоров за соответствующий месяц

**0 1 2 3 4 …** - кол-во месяцев, прошедших с момента заключения договора (MOB – Month on Book)

На пересечении поколения и MOB находится сумма по задолженности 90+ среди всех договоров из соответствующего поколения и месяца жизни кредита.  
  
Решением является файл .excel со всеми рассчетами.

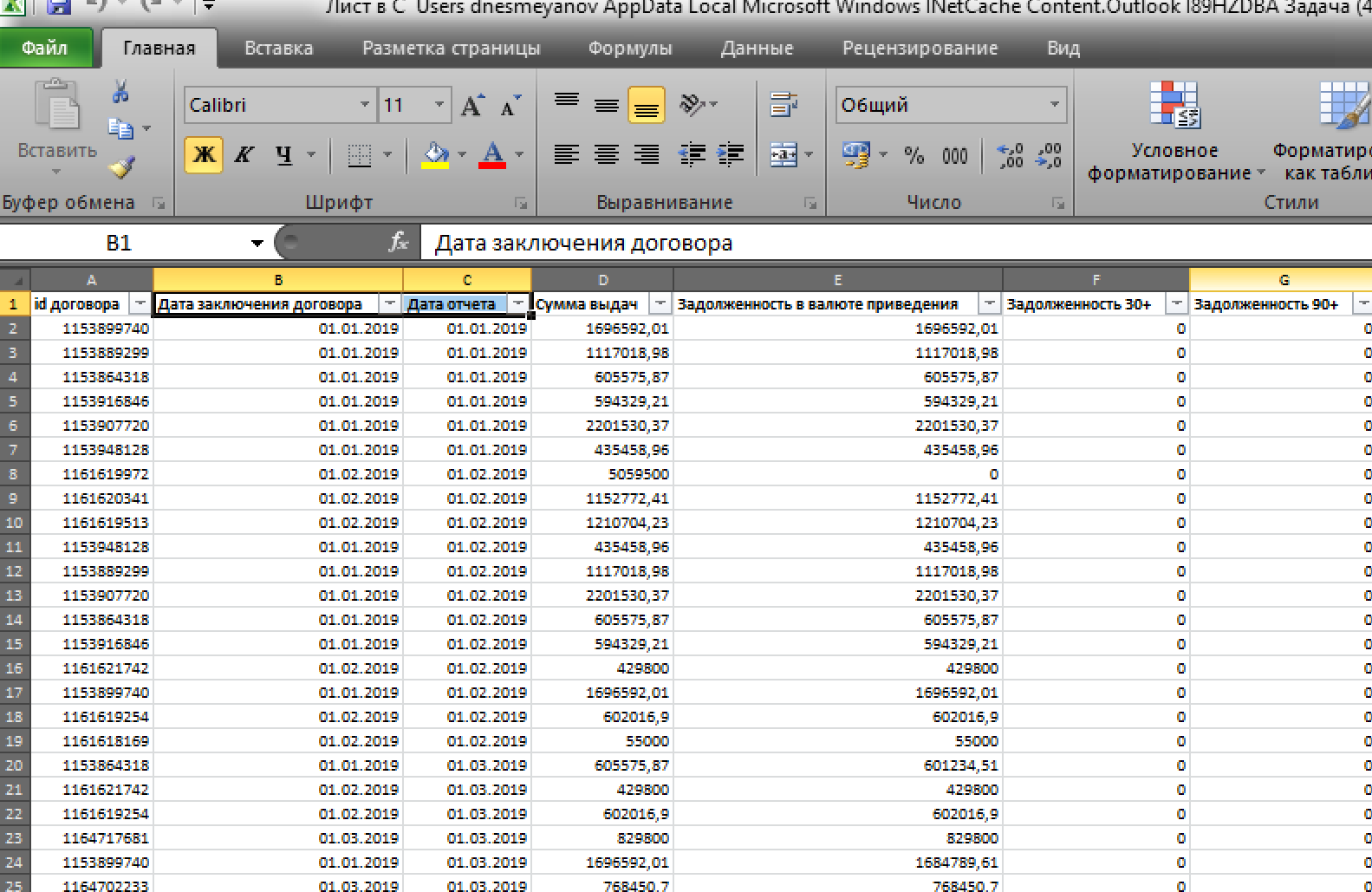
Подсказка:  


Рисунок 2 – Данные из «Пример винтаж»

Количеством месяцев с момента заключения договора (MOB) будет являться Разница дат (в месяцах) между Столбцами «Дата заключения договора» и «Дата отчета» (Рисунок 2).

**Задача 2:**

Представьте, что данные из файла (Пример «Винтаж») хранятся в таблице базы данных SQL (sql\_db.dbo.table\_vintaj)  
Напишите скрипт на языке SQL, который для каждого поколения выводил бы количество договоров, количество договоров с просрочкой в 90+ и результат деления первого числа на второе (% простроченных договоров) (Таблица 1).  
  
Решением является файл с разрешением .sql/

Таблица 1

|  |  |  |  |
| --- | --- | --- | --- |
| Поколение | Кол-во договоров | Кол-во 90+ | % Дефолтов |
| 01.01.2019 | 10 | 1 | 0,1 |
| 01.02.2019 | 12 | 0 | 0 |
| … | … | … | … |

**Задание повышенной сложности\* (Будет преимуществом)**Добавьте к запросу столбец, в котором бы каждому поколению присваивалась цифра в обратном порядке, где у самого последнего поколения будет 1, а у самого древнего n, где n – количество поколений (Таблица 2).

Таблица 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Поколение | Кол-во договоров | Кол-во 90+ | % Дефолтов | Номер |
| 01.01.2019 | 10 | 1 | 0,1 | 6 |
| 01.02.2019 | 12 | 0 | 0 | 5 |
| 01.03.2019 | … | … | … | 4 |
| 01.04.2019 |  |  |  | 3 |
| 01.05.2019 |  |  |  | 2 |
| 01.06.2019 |  |  |  | 1 |

**Задача 3:**

С помощью Python необходимо написать функцию, которая получает на входе винтаж: pandas.DataFrame и возвращает винтаж с просроченной задолженностью в процентах от Суммы выдач: pd.DataFrame.   
Например (см. Рисунок 1), на шестом MOB поколения 01.09.2019 сумма просроченной задолженности составила 2 051 029, Сумма выдач в этот месяц составила 208 996 772. Значит, в процентах от суммы выдач на 6 MOB просрочено 2 051 029/208 996 772 ~ 0,01 (или 1%)  
  
Решением является файл с функцией .ipynb или .py/

-------  
1) Можно использовать другой фреймворк (не pandas);  
2) Постарайтесь максимально оптимизировать вычисления (использовать матричную алгебру).

**Задача 4:**

На основе предложенных данных постройте модель бинарной классификации.   
  
Рассчитайте метрики, подготовьте графики.   
Решением будет файл .ipynb с вашими расчётами и графиками.

------

1. Модели можно использовать любые, логистическая регрессия тоже хороший вариант;
2. Объясните в ноутбуке, почему вы использовали именно данную метрику.

------------------------------------------------

Архивируйте все файлы в одной папке .rar/  
Все должно работать «из архива».  
\* Использовать google collab можно, но постарайтесь, чтобы код был воспроизводимый.