ЛШ РАИИ 2024 (г. Коломна)

Задание на хакатон

# Описание задания

"Задача нового этапа в горизонте текущего десятилетия - обеспечить именно массовое внедрение искусственного интеллекта, оно должно охватить все отрасли экономики и социальной сферы и систему госуправления", - указал Президент Российской Федерации В.В. Путин на международной конференции "Путешествие в мир искусственного интеллекта" в ноябре 2022 года [1].

Рассмотрим непосредственно ту сферу, которая касается каждого из нас – потребление коммунальных ресурсов: холодной, горячей воды и электроэнергии.

Как известно, за потребление ресурсов необходимо платить ежемесячно, причём каждый месяц сумма оплаты отличается. Такая сумма оплаты в отрасли называется «**начисление**» [2], т.к. речь идёт обычно о долгах за потребление. Для каждой квартиры заводится «**лицевой счёт**», и собственники помещений должны следить за ежемесячными начислениями и производить «**оплату**». Обычно оплата производится за предыдущий месяц. Иногда у людей копится «**задолженность**» на лицевом счёте, когда оплата ниже начислений.

Например, счастливый собственник приобрёл квартиру. Ему завели лицевой счёт с балансом *0 рублей*. В первый месяц житель потратил ресурс «электричество» на *1000 рублей*. Таким образом, на начало второго месяца долг будет составлять 1000 рублей (то есть баланс *-1000 рублей*), оплачено в первом месяце было *0 рублей*. Допустим, что во втором месяце собственник заплатил *1500 рублей* и потратил на электричество ещё *700 рублей*. Таким образом, на начало третьего месяца долг будет составлять рублей.

На основании примера можно сделать вывод, что при адекватном паттерне оплаты житель всегда будет задолжать ресурсоснабжающей компании примерно за один период (месяц), что является нормальным для отрасли.

Но, к сожалению, не все жители платят за коммунальные ресурсы вовремя и в полном объёме. Из-за этого бывает сложно спрогнозировать «**собираемость денежных средств**», т.е. отношение .

Компания, которая занимается расчётом начислений и оплат одного региона, предоставила анонимные данные о потреблении коммунальных ресурсов в течение года, а также долгов, начислений и оплат, и просит (задачи):

1. Выявить возможные паттерны клиента-плательщика.
2. Определить реальное количество проживающих на основании потребления коммунальных ресурсов.
3. Предложить возможные применения предоставленных данных для бизнеса.

# Описание набора данных

Для анализа и прогнозирования предоставляется архив *raai\_school\_2024.zip* в формате ZIP с единственным файлом *raai\_school\_2024.csv* в формате CSV (разделители - «;», заголовок в первой строке).

Список атрибутов:

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Название** | **Тип** | **Описание** |
| 1 | house\_tkn | int | ИД (токен) дома. |
| 2 | flat\_tkn | int | ИД (токен) квартиры. |
| 3 | payment\_period | ГГГГ-ММ-01 | Период оплаты (1 день месяца). |
| 4 | income | decimal | Начислено в периоде. |
| 5 | debt | decimal | Долг на начало периода с учётом оплаты. |
| 6 | raised | decimal | Оплачено в периоде. |
| 7 | volume\_cold | float | Объём потребления ХВС, . |
| 8 | volume\_hot | float | Объём потребления ГВС, . |
| 9 | volume\_electr | float | Объём потребления Электроэнергии, квт\*ч. |

Количество строк в наборе данных: 3206079.

Примечания:

* Положительная сумма в поле *debt* – это долг, отрицательная – переплата.
* Некоторые периоды могут отсутствовать в наборе данных.

# Оформление результата

На основании анализа набора данных должна быть составлена презентация (для задач 1 и 3) и предоставлен результирующий файл (для задачи 2).

Описание результирующего файла для проверки:

* Формат файла: CSV, аналогичный заданию.
* Список атрибутов: flat\_tkn (int), residents\_avg\_count (float), где:
  + flat\_tkn – ИД (токен) квартиры;
  + residents\_avg\_count – Среднее количество проживающих в квартире.

# Список литературы

1. <https://tass.ru/ekonomika/16418761>
2. <https://www.aisgorod.ru/products/details/system_charges/>