

SQL project

Анализ набора данных ресторанов с оценкой клиентов

by Leysan Iskhakova, ML

Набор данных: Restaurant Data with Consumer Ratings

Источник: [kaggle.com](https://www.kaggle.com)

Описание

Набор данных состоит из 9 файлов в формате csv: 5 файлов с данными о ресторанах, 4 файла с данными о клиентах (пользователях) и один файл с оценками ресторанов пользователями

Цель

Проанализировать датасеты и составить рейтинг ресторанов на основе оценок пользователей с помощью SQL

Первичная обработка данных

Инструменты: Python, Jupyter notebook

Библиотеки: pandas, ydata_profiling, sqlite3

Файлы: restaurant_analysis.ipynb и отчеты анализа датасетов в формате html

**согласно отчетам в во всех
датафреймах нет
отсутствующих, пустых значений
и дубликатов**

**названия колонок были
переименованы согласно Naming
convention: латинские буквы,
нижний регистр, разделитель _**

Overview	Alerts 3	Reproduction
Dataset statistics		
Number of variables	5	
Number of observations	1161	
Missing cells	0	
Missing cells (%)	0.0%	
Duplicate rows	0	
Duplicate rows (%)	0.0%	
Total size in memory	45.5 KiB	
Average record size in memory	40.1 B	

Пример из отчета датайрема rating_final

Загрузка данных в базу данных

База данных: restaurant_rating.db в SQLite

Таблицы: 9 таблиц из соответствующих датафреймов

В SQLite нельзя определять ключи с помощью инструкция ALTER TABLE. Это можно сделать только в первоначальной инструкции CREATE TABLE

**Создание таблиц с ключами и
их наполнение данными из
старых таблиц**

```
•-- rating definition
ALTER TABLE rating RENAME TO old_rating;
•CREATE TABLE rating (
    line_number INTEGER,
    user_id TEXT,
    place_id INTEGER,
    rating INTEGER,
    food_rating INTEGER,
    service_rating INTEGER,
    CONSTRAINT line_number_r PRIMARY KEY (line_number),
    FOREIGN KEY(user_id) REFERENCES consumer_info(user_id),
    FOREIGN KEY(place_id) REFERENCES restaurant_info(place_id)
);
INSERT INTO rating SELECT * FROM old_rating;
```

Анализ информации о ресторанах

Количество ресторанов - 130

```
SELECT COUNT(place_id) AS count_rest  
FROM restaurant_info
```

*Количество ресторанов по
ценовому сегменту*

price	count_rest
medium	60
low	45
high	25

Количество ресторанов с алкоголем

alcohol	count_rest
No_Alcohol_Served	87
Wine-Beer	34
Full_Bar	9

Анализ информации о ресторанах

Виды кухни ресторанов

```
SELECT  
rc.r_cuisine AS cuisine,  
COUNT(rc.r_cuisine) AS count_rest  
FROM restaurant_info ri  
LEFT JOIN restaurant_cuisine rc |  
ON ri.place_id = rc.place_id  
GROUP BY cuisine  
ORDER BY count_rest DESC
```

	cuisine	count_rest
1	Mexican	28
2	Bar	13
3	Cafeteria	9
4	Fast_Food	8
5	Bar_Pub_Brewery	6

Анализ информации о ресторанах

Принимаемые способы оплаты в ресторанах

```
SELECT
ROUND(AVG(count_payment)) AS avg_count_payment,
MAX(count_payment) AS max_count_payment,
MIN(count_payment) AS min_count_payment
FROM (
    SELECT
        ri.place_id,
        COUNT(rp.r_payment) OVER (PARTITION BY rp.place_id AS count_payment
    FROM restaurant_info ri
    LEFT JOIN restaurant_payment rp ON rp.place_id = ri.place_id
    ORDER BY count_payment DESC
)
WHERE count_payment > 0
```

	payment
1	American_Express
2	Carte_Blanche
3	MasterCard-Eurocard
4	VISA
5	bank_debit_cards
6	cash

	payment
1	cash

Анализ информации о пользователях

Количество пользователей - 138

```
SELECT COUNT(user_id)
FROM consumer_info
```

Бюджет

	ASC budget	123 count_us
1	medium	91
2	low	35
3	?	7
4	high	5

Употребление алкоголя

	ASC drink_level	123 count_us
1	abstemious	51
2	casual drinker	47
3	social drinker	40

Анализ информации о пользователях

Предпочитаемые виды кухни пользователей

```
SELECT  
cc.u_cuisine AS cuisine,  
COUNT(ci.user_id) AS count_user  
FROM consumer_info ci  
LEFT JOIN consumer_cuisine cc  
ON cc.user_id = ci.user_id  
GROUP BY cc.u_cuisine  
ORDER BY count_user DESC
```

	cuisine	count_user
1	Mexican	97
2	American	11
3	Pizzeria	9
4	Cafeteria	9
5	Family	8
6	Cafe-Coffee_Shop	8
7	Japanese	7
8	Italian	7
9	Latin_American	6
10	Hot_Dogs	6

Анализ информации о пользователях

Предпочитаемые способы оплаты пользователей

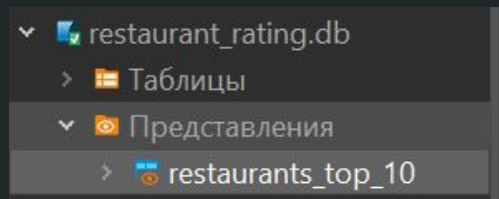
```
SELECT
  cp.u_payment AS payment,
  COUNT(ci.user_id) AS count_user
FROM consumer_info ci
LEFT JOIN consumer_payment cp
ON cp.user_id = ci.user_id
GROUP BY payment
ORDER BY count_user DESC
```

	asc payment	123 count_user	
1	cash	131	
2	bank_debit_cards	22	
3	VISA	17	
4	[NULL]	5	
5	MasterCard-Eurocard	4	
6	American_Express	3	

Рейтинг ресторанов по оценке пользователей

	123 place_id ▾	ABC name ▾	123 overall_rating ▾	123 food_rating ▾	123 service_rating ▾
1	135 085	Tortas Locas Hipocampo	48	53	42
2	132 825	puesto de tacos	41	43	30
3	135 032	Cafeteria y Restaurant El Pacifico	33	30	30
4	135 052	La Cantina Restaurante	32	29	29
5	135 038	Restaurant la Chalita	29	28	23
6	135 060	Restaurante Marisco Sam	25	23	28
7	135 025	El Rincon de San Francisco	25	22	24
8	132 862	La Posada del Virrey	25	25	26
9	132 834	Gorditas Doa Gloria	25	28	22
10	135 062	Restaurante El Cielo Potosino	26	26	23

Сохранен как VIEW (представление)



Анализ рейтинга ресторанов по оценке пользователей

Виды кухни ресторанов в топ 10 рейтинга

```
SELECT *  
FROM (  
  SELECT  
    rc.r_cuisine AS cuisine,  
    COUNT(rt.place_id) OVER (PARTITION BY r_cuisine) AS count_top_rest  
  FROM restaurants_top_10 rt  
  LEFT JOIN restaurant_cuisine rc ON rc.place_id = rt.place_id  
)  
GROUP BY cuisine  
ORDER BY count_top_rest DESC  
|
```

	cuisine	count_top_rest
1	Mexican	3
2	[NULL]	2
3	Seafood	1
4	International	1
5	Fast_Food	1
6	Contemporary	1
7	Cafeteria	1
8	Bar_Pub_Brewery	1
9	Bar	1

Анализ рейтинга ресторанов по оценке пользователей

*Ценовой сегмент ресторанов
в топ 10 рейтинга*

	price	count_top_rest
1	medium	6
2	low	2
3	high	2

*Продажа алкоголя в ресторанах
в топ 10 рейтинга*

	alcohol	count_top_rest
1	No_Alcohol_Served	7
2	Wine-Beer	2
3	Full_Bar	1

*Количество способов оплаты в
ресторанах в топ 10 рейтинга*

	name	count_payment
1	Tortas Locas Hipocampo	1
2	puesto de tacos	1
3	Cafeteria y Restaurant El Pacifico	3
4	La Cantina Restaurante	3
5	Restaurant la Chalita	1
6	Restaurante Marisco Sam	1
7	La Posada del Virrey	5
8	Restaurante El Cielo Potosino	1
9	Gorditas Doa Gloria	1
10	El Rincon de San Francisco	4

Выполнено

- проанализированы датасеты
- составлен рейтинг ресторанов на основе оценок пользователей и тоже проанализирован

Выводы

Согласно набору данных, люди в Мексике предпочитают мексиканскую кухню и оплату наличными

Планы на будущее

- проанализировать какие факторы влияют на рейтинг ресторана
- создать рекомендательную систему по ресторанам: предлагать пользователю рестораны в соответствии с его предпочтениями

Спасибо за внимание!
