Демонстрационный вариант 2022/2023. Задачный тур заключительного этапа

Время выполнения – 240 минут

Максимальное количество баллов – 100

Задание 1. – 25 баллов

Перед вами текст презентации одной сети пиццерий для потенциальных инвесторов. Вы являетесь аналитиком инвестиционной компании и хотите произвести оценку данной компании по информации ниже. Аналитик, который готовил презентацию, использовал различные ухищрения, чтобы выставить компанию в более выгодном свете. Вам нужно их найти. Выпишите все проблемы и обоснуйте их, при надобности приведите поясняющие примеры.

«Фэмили Пицца» является современной сетью пиццерий семейного формата с широким ассортиментом. Пока сеть состоит из двух пиццерий в районе Девятково, а в прошлом месяце открылась еще одна в районе Десятково.

На рис. 1 продемонстрирован стремительный рост заказов в нашей сети за последние полгода.

Рисунок 1

А при анализе рынков в этих двух районах мы выяснили, что за прошлый месяц доли рынков в каждом из этих районов выросли.

Главным образом пиццерия ориентирована на семейных клиентов с детьми дошкольного возраста – специально для этого разработано детское меню, а в каждой пиццерии есть зона для детей с различными играми. Ниже на рис. 2 Вы можете заметить стабильность количества посетителей с детьми возраста 0-7 лет.

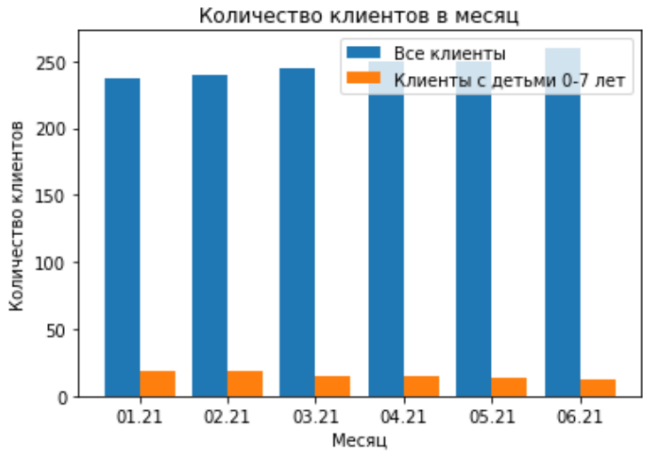

Рисунок 2

Наша компания заботится и о будущем подрастающего поколения, поэтому мы пытаемся свести количество отходов к минимуму и используем вторсырье по возможности. На рис. 3 вы можете заметить, что в нашей сети используется минимальное количество неперерабатываемого пластика на одного посетителя. Более того, наша компания с гордостью носит титул ТОП-1 среди экологичных пиццерий района.

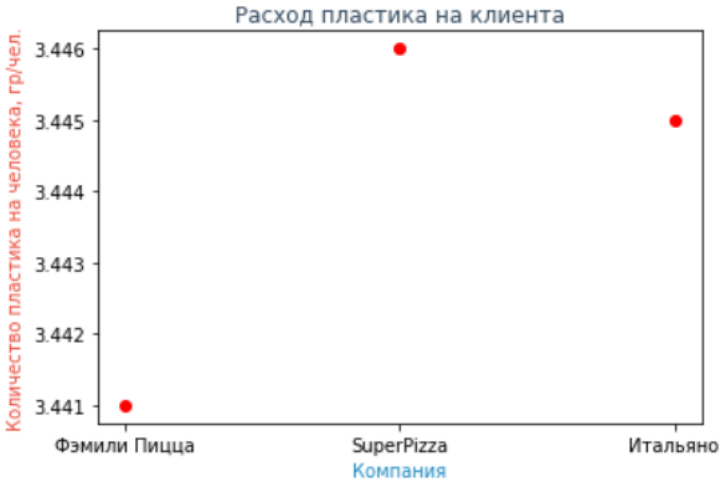

Рисунок 3

Посетители высоко оценивают качество блюд и обслуживания: на платформе Яндекс.Карты средняя оценка нашей сети 4.9 (у главных конкурентов 4.7). В условиях пандемии мы так же разработали интернет-платформу для заказов. После каждого заказа мы проводим опрос по удобству сайта, средняя оценка пользователей 4.96, что выше средних оценок аналогичных сервисов конкурентов в Google Play / AppStore.

Исходя из данных выше, мы имеем стабильное конкурирующее положение по рынку, а по некоторым показателям уже превосходим конкурентов.

Задание 2. – 25 баллов

Петя работает в маркетинговом отделе и занимается подготовкой данных для рубрики «интересные факты о вас», которая отображается клиентам. За прошлую неделю Петя подготовил много выборок, но у одной из потерялось описание. Эта выборка дана в файле target.csv. Однако Петя помнит, что для расчетов ему понадобился в том числе граф связей клиентов друг с другом. Этот граф лежит в файле graph_example.csv

Помогите Пете – объясните, как строилась выборка и по какому принципу проставлялся target. Вам не даны все данные, которые использовал Петя, поэтому у вас может не получиться повторить target идеально. Однако вы можете искать и подтверждать зависимости, которые Петя использовал. Воссоздайте искомую величину и проверьте ее корреляцию с исходным target. Заведомо известно, что можно получить корреляцию > 0.99.

Описание полей graph_example.csv:

- v1 – ID первого клиента в связи
- v2 – ID второго клиента в связи
- weight – «вес» связи, ее значимость

Описание полей target.csv:

- vertex – ID клиента
- target – некоторая величина, характеризующая клиента, описание которой требуется восстановить

Задание 3. – 25 баллов

Петя решил исследовать статистику трат клиентов банка в зависимости от региона. Для этого он собрал выборку, содержащуюся в файле task3_data.csv. Данные описываются следующими полями:

- party_rk – идентификатор клиента
- purchase_sum – сумма трат за определенный период
- russian_region_nm – название региона проживания клиента
- russian_region_cd – код региона проживания клиента
- timediff_to_msk_hour_cnt – разница в часах между часовым поясом проживания клиента и МСК
- russian_federal_district_cd – код федерального округа проживания клиента

Петю интересуют следующие вопросы о выборке:

1. Как различаются траты населения в разных частях страны?
2. Какие регионы самые «тратящие»? Какие самые богатые?
3. Где наибольший разрыв в доходах?
4. Влияет ли часовой пояс на то, сколько тратят люди?

Попробуйте ответить на эти вопросы с помощью визуализации. Будьте лаконичны – максимум 3 графика. Не используйте в визуализации карту, это не требуется и не приведет к получению дополнительных баллов. Прокомментируйте графики и сделайте выводы, где возможно. Если вы делали предобработку данных, то объясните ее и приложите файл.

Задание 4. – 25 баллов

Вам дан датасет с 6 признаками. 4 из этих признака были получены синтетически из других двух базовых с помощью различных математических операций с добавлением некоторого шума. Найдите базовые признаки и алгоритм получения синтетических признаков. Опишите математические преобразования, с помощью которых получились сгенерированные признаки. Главное описать тип преобразований, не указывая точные коэффициенты. Необходимые данные находятся в файле task4_data.csv