Итоговый проект: Uber & Lyft.

Summary:

Задача - на тестовом наборе данных от Uber & Lyft, содержащем множество параметров, провести обработку данных и настройку модели, построение связей в модели данных, воспользоваться DAX-формулами и построить графики. Визуализация с помощью Python, загрузка отчета.

Файл для работы по ссылке:

https://hb.bizmrg.com/productstar\_py-bi-bd/Power%20BI/8.%20%D0%98%D1%82%D0%BE%D0%B3%D0%BE%D0%B2%D1%8B%D0%B9%20%D0%BF%D1%80%D0%BE%D0%B5%D0%BA%D1%82/rideshare\_kaggle.csv

Столбцы, которые необходимо оставить в модели:

1. cab\_type – тип такси (Uber или Lyft);2. cloudCover – облачность;3. datetime – дата и время;4. destination – пункт назначения;5. distance – расстояние между точкой отправления и точкой назначения;6. humidity – влажность;7. latitude – широта;8. longitude – долгота;9. moonPhase – фаза Луны;10. name – название тарифа;11. pressure – давление;12. price – цена поездки;13. short\_summary – краткое описание погоды (текст);14. source – пункт отправления;15. surge\_multiplier – мультипликатор цены из-за повышенного спроса;16. temperature – температура;17. visibility – видимость;18. windSpeed – скорость ветра.

Этапы проекта:

1. Подготовительный этап по обработке данных и настройке модели. Удалены лишние столбцы, оставив только те, что указаны в описании проекта. Созданы отдельные таблицы-справочники: source, destination, cab\_type, name.

Создана отдельная таблица календаря, которая содержит (Даты с 1 ноября по 31 декабря 2018 года, Номер дня недели). В модели данных таблицы-справочники и таблицу календаря соединены с основным набором данных.

2. Добавлен параметр, который принимает значения от 1 до 3 шагом 0.25, и доработана формула рассчитывающая долю поездок выше выбранного параметра. Добавлен срез с параметром на лист отчета и карточка с доработанной формулой.

3.Создание визуализаций.

Построена столбчатая диаграмма (гистограмма) которая визуализирует суммарные стоимости поездок (price) по дням (date).

Дополнительная визуализация типа “WordCloud” - диаграмма использует столбец short\_summary, описывающий погоду.

Тип визуализации “карта”отображает координаты каждой поездки.

Иерархия отражающая зависимость количества поездок от дня недели и часа.

Вывод: количество поездок не зависит от часа дня и одинаково утром, днем, вечером и ночью.

4. Визуализация с помощью Python. Построена визуализация в python для пункта назначения North Station, использующая поля: price, cloudCover, distance, moonPhase, humidity, pressure, surge\_multiplier, temperature, visibility, windSpeed – изучены характеристики которые влияют на поездки.

5. Построение DAX-формул.

В этой части итогового проекта созданы несколько DAX-формул и построены графики, отвечающие на вопросы. Например, создана мера, которая подсчитывает общее количество поездок.