**1. Бизнес-анализ (Business Understanding)**

В первую очередь необходимо определиться с целями и скоупом проекта.

Для этого нужно найти ответы на следующие вопросы:

* *Организационная структура: кто участвует в проекте со стороны заказчика сервиса, кто будет основным пользователем?*

Проект реализует один разработчик в качестве итогового проекта на образовательных курсах. Основными заинтересованными сторонами данного решения могут выступать городские органы управления движением или частный предприниматель, реализующий урбанистический проект. Косвенными выгодополучателями проекта также выступят участники дорожного движения, которым станет удобнее передвигаться по городу

* *Какова бизнес-цель проекта? Например, уменьшение оттока клиентов.*

Главная бизнес-цель проекта – снижение стоимости такси и каршеринга, повышение его доступности и снижение простоев. а

* *Существуют ли какие-то уже разработанные решения? Если существуют, то какие и чем именно текущее решение не устраивает?*

Крупные сервисы такси и каршеринга (Uber, Lyft, DiDi и др.) используют собственные системы прогнозирования спроса и динамического ценообразования.

**1.1 Текущая ситуация (Assessing current solution)**

Оцениваем, хватает ли ресурсов для проекта.

* Есть ли доступное железо или его необходимо закупать?

Разработка будет вестись на железе MacBook Air M4

* Где и как хранятся данные, будет ли предоставлен доступ в эти системы, нужно ли дополнительно докупать/собирать

Будут использоваться данные низ открытых источников:

1. NYC TLC Trip Record Data– исторические данные о поездках такси
2. Uber movement – средние значения времени пути по районам
3. OpenStreetMap – геоднанные по районам
4. NOAA Weather Data – погодные условия
5. Какие-то данные о концертах и массовых мероприятиях
6. Данные о демографии в районах

**Риски:**

|  |  |  |
| --- | --- | --- |
| **Название риска** | **Суть** | **Способ митигации** |
| Ограничения по ресурсам | Проект выполняется одним студентом с ограниченными проихводственными мощностями | Использование дополнительных вычислительных мощностей и простых моделей |
| Проблемы с данными | Данные по NYC огромный и могут содержать ошибки, данные Uber содержат только среднее время поездок | Использование датасетов с другими признаками, чтобы совместить с текущими |

**1.2 Решаемые задачи с точки зрения аналитики (Data Mining goals)**

Выполняем постановку в технических терминах. Для этого нужно ответить на следующие вопросы:

Необходимо будет прогнозировать количество поездок по району и времени, это регрессионная задача.

В таком случае для оценки модели отлично подойдет RMSE и R2.

R2 должен быть выше 0,7; RMSE выше чем у константного прогноза.

**1.3 План проекта (Project Plan)**

Бизнес-понимание – 2 июня

Изучение данных – 3–12 июня

Подготовка данных 12–25 июня

Моделирование 25-11 июля

Оценка 12-16 июля