

Разработка концепции работы с большими данными  
по прогнозированию потребительской активности  
в Международном аэропорту Шереметьево



**ШЕРЕМЕТЬЕВО**  
МЕЖДУНАРОДНЫЙ АЭРОПОРТ

Работу выполнил  
Дробященко Игорь



# Оглавление

Об отчете	2
1 Описание решения	3
1.1 Описание концепции и подхода к работе с данными	4
1.2 Обоснование выбора используемых данных	5
1.3 Описание гипотез	6
1.4 Описание параметров	8
1.5 Прогноз влияния на выручку	9
1.5 Возможные управленческие решения	10
2. Итоги	11



# Описание решения





## Описание концепции и подхода к работе с данными

Задача: спрогнозировать выручку торговых точек в Шереметьево

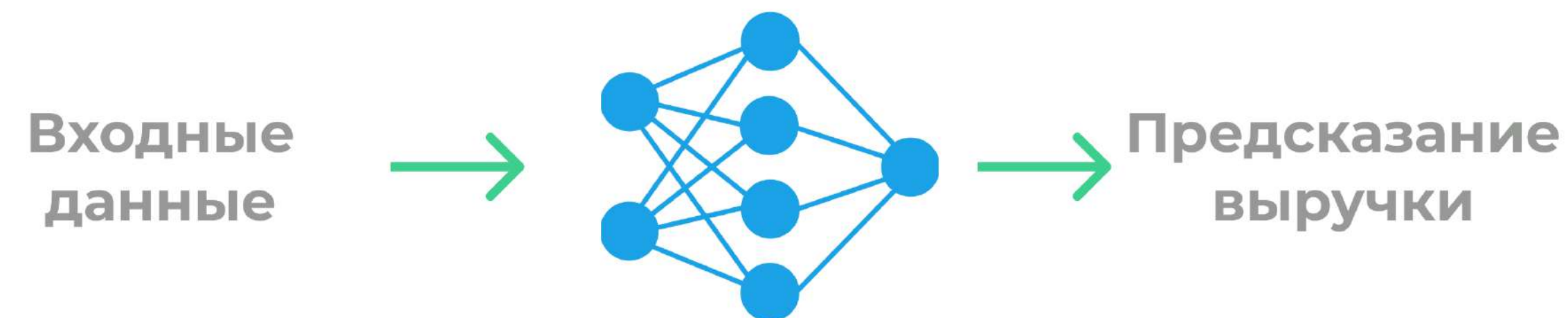
Концепция:

1. Собираем все данные в одну таблицу
2. Получаем данные из внешних источников
3. Создаем новые признаки
4. Строим мат. модель
5. Оцениваем точность мат. модели
6. Оцениваем влияние признаков на доход
7. Моделируем ситуацию в будущем

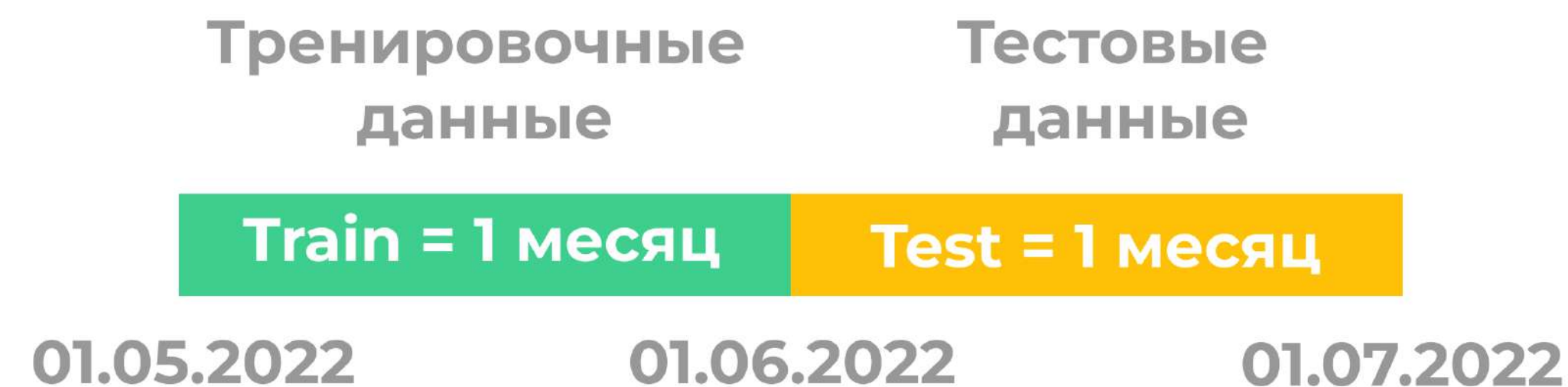
Используемые данные:

1. Расписание прилетов и вылетов
2. Погода METAR
3. Онлайн табло а/к Аэрофлот и Россия
4. Список аэропортов и самолетов
5. Выручка торговых точек

### Принцип работы ML алгоритмов



### Принцип обучения моделей



### Пример данных для обучения ML модели

№ точки	Дата и время	Всего рах	AFL рах	SDM рах	C class рах	Y class рах	B term рах	...	Доход
5	04.05.2022 14:30	2100	2100	500	130	2100	2000	...	5000
23	06.05.2022 16:00	3450	3450	700	120	3450	1500	...	3250
16	11.05.2022 07:30	1700	1700	1100	200	1700	3000	...	2750
30	20.05.2022 14:30	4000	4000	1000	150	4000	4250	...	9400



# Обоснование выбора используемых данных

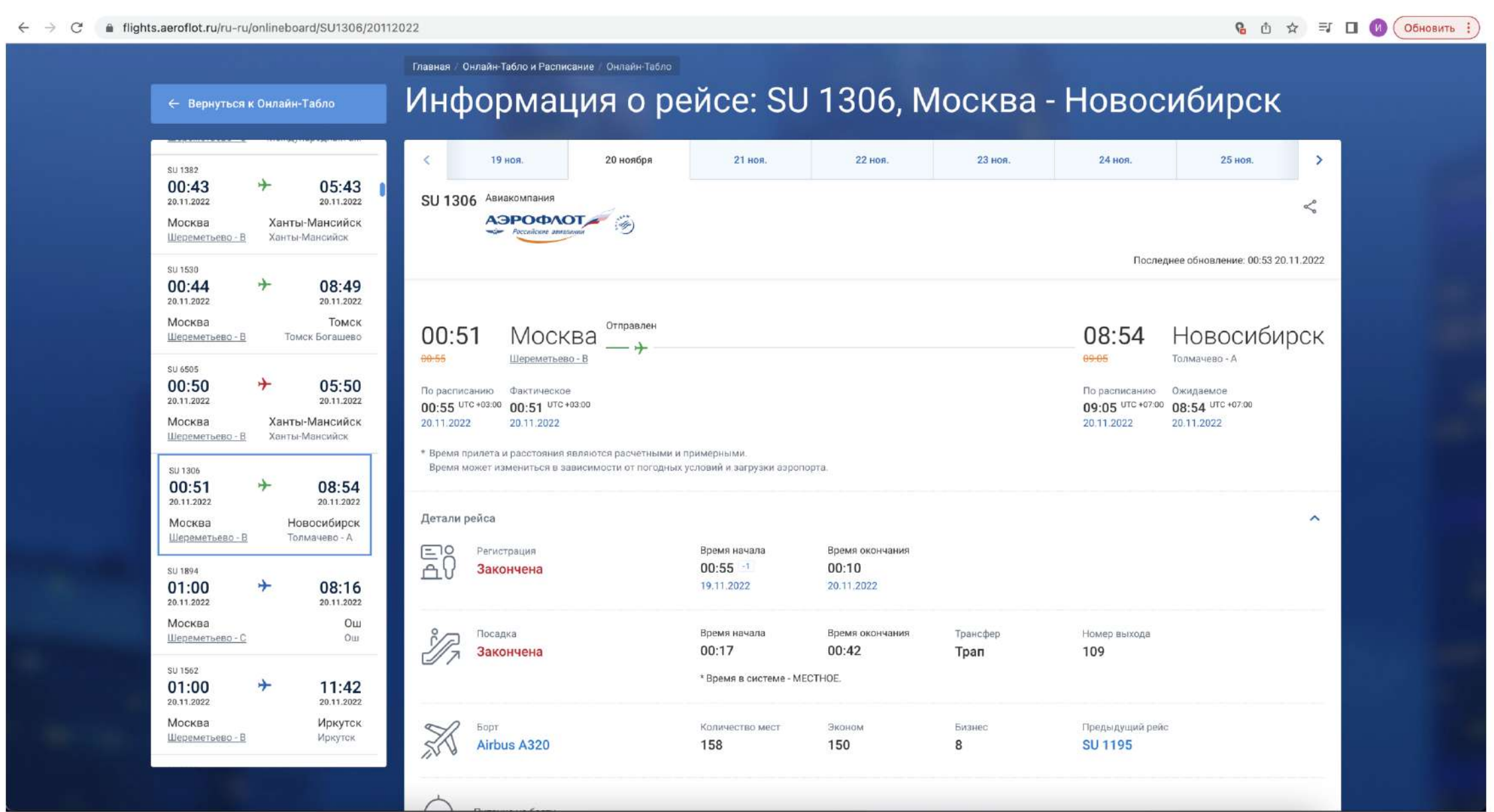
## Используемые данные:

- 1. Расписание вылетов и прилетов
- 2. Погода METAR. В зависимости от погоды может меняться выручка магазинов.
- 3. Онлайн табло а/к Аэрофлот и Россия. Много дополнительных данных: номер гейта, время задержки рейсов, время начала и окончания посадки, кол-во мест бизнес и эконом класса и тд. Аэрофлот+Россия > 83% рейсов из/в SVO
- 4. Список аэропортов, содержащий их координаты по широте и долготе, а также страну, коды icao и iata. С их помощью можно вычислить длительность полета, направление (юг или восток), страну прибытия
- 5. Список самолетов и их компоновка (количество мест в бизнес и эконом классе)
- 6. Выручка торговых точек. Целевые данные.

## Пример данных для обучения ML модели

№ точки	Дата и время	Всего рах	AFL рах	SDM рах	C class рах	Y class рах	B term рах	...	Доход
5	04.05.2022 14:30	2100	2100	500	130	2100	2000	...	5000
23	06.05.2022 16:00	3450	3450	700	120	3450	1500	...	3250
16	11.05.2022 07:30	1700	1700	1100	200	1700	3000	...	2750
30	20.05.2022 14:30	4000	4000	1000	150	4000	4250	...	9400

## Онлайн табло Аэрофлота и России





# 1.3

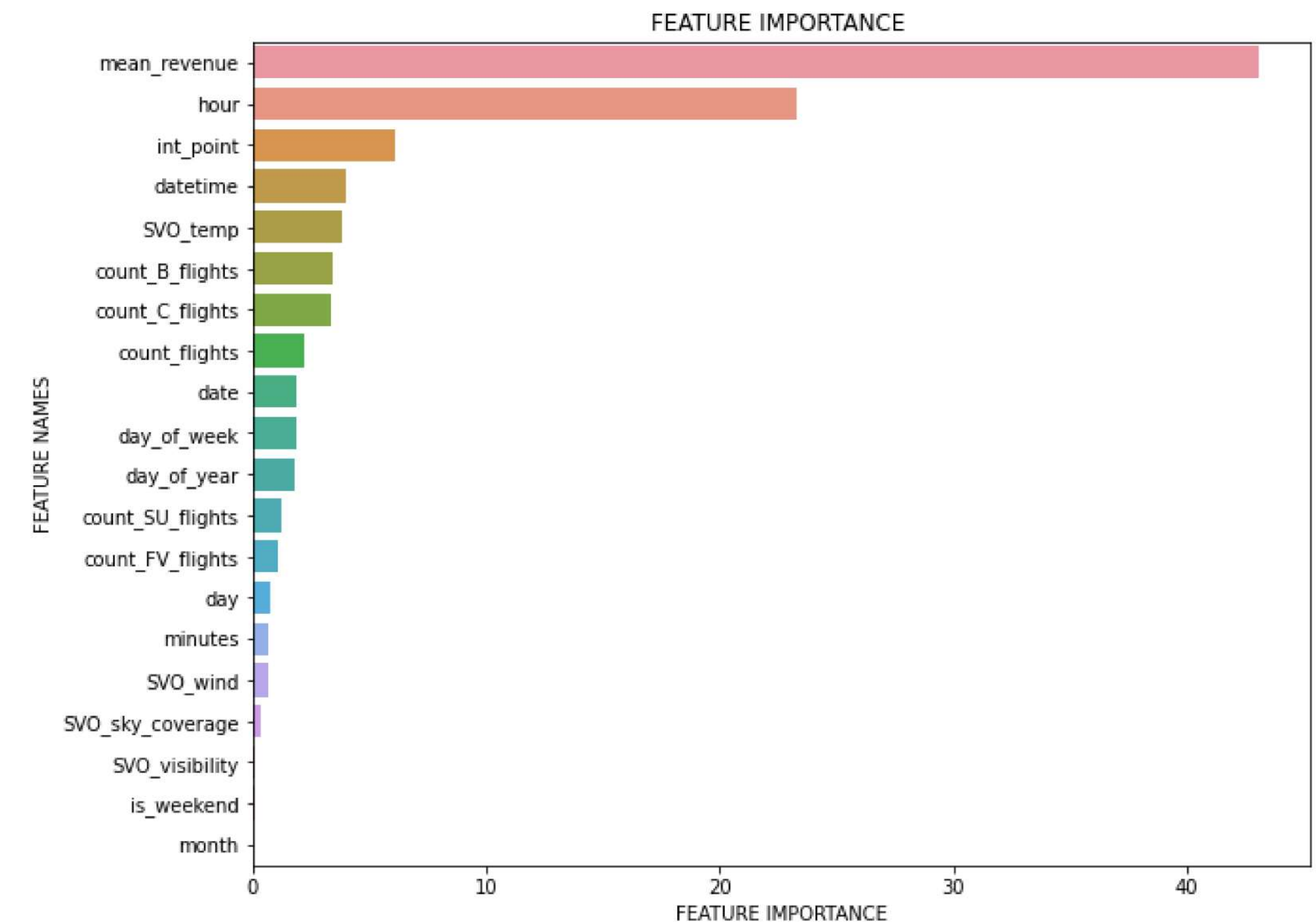
## Описание гипотез

Ключевые гипотезы, которые “сработали”:

1. Прямая зависимость между количеством пассажиров в чистой зоне в выручке торговых точек
2. У каждой точки есть свой baseline: уровень выручки при наборе внешних факторов
3. Сильная зависимость выручки от времени суток
4. потребительская активность пассажиров зависит от авиакомпании, класса обслуживания (С/У), погоды в SVO и городе прилета (является ли курортом, расположен севернее/южнее/восточнее Москвы) и ряда других параметров рейса
5. У каждой торговой точки свой портрет целевой клиента
6. Пассажиры покупают преимущественно рядом со своими гейтами

Если строить единую модель для всех точек, то ключевой признак - средняя выручка точки.

Далее, в зависимости от времени и др. критериев, прогноз выручки корректируется





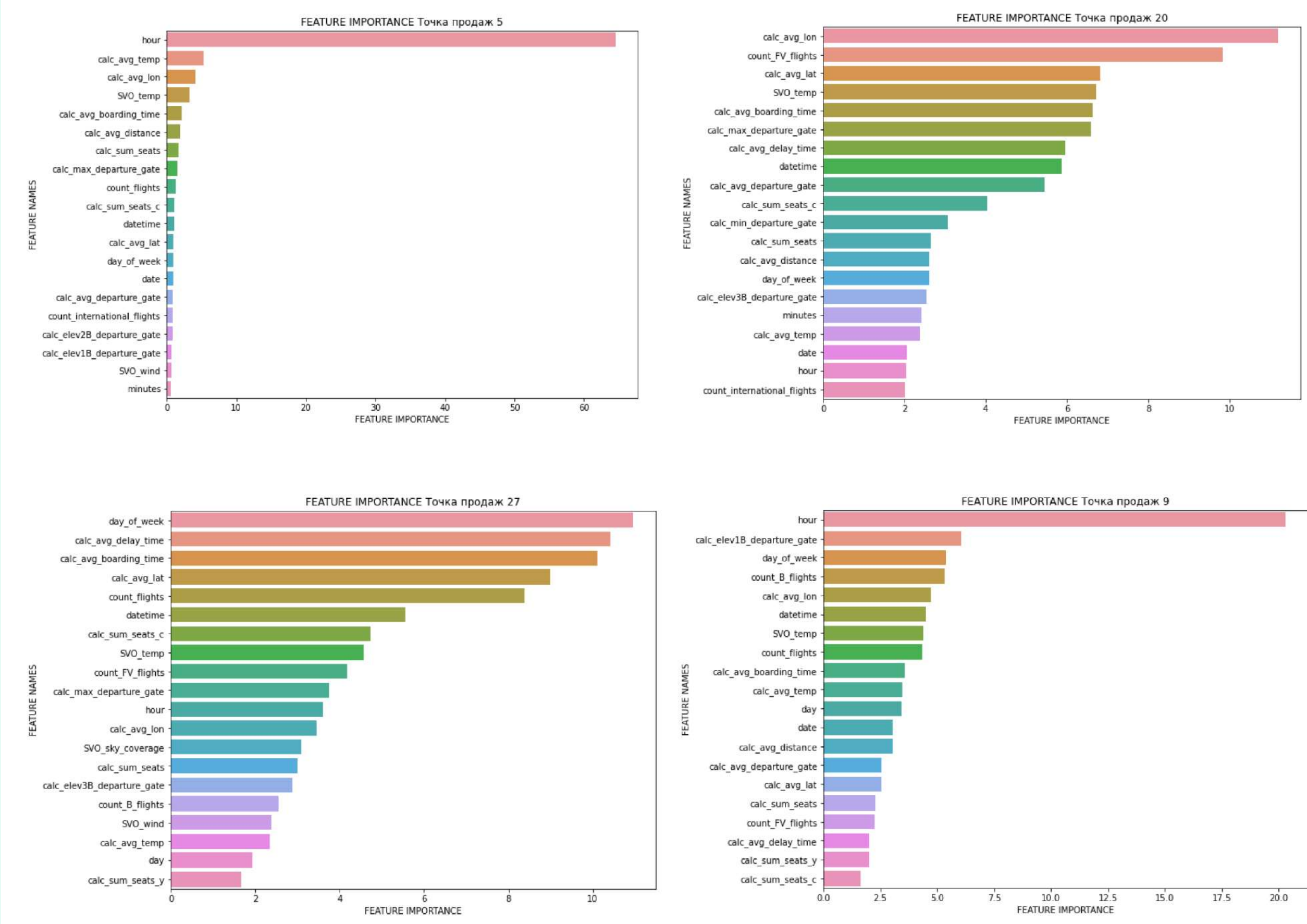
# Описание гипотез

## Ключевые гипотезы, которые “сработали”:

- 1. Прямая зависимость между количеством пассажиров в чистой зоне в выручке торговых точек
- 2. У каждой точки есть свой baseline: уровень выручки при наборе внешних факторов
- 3. Сильная зависимость выручки от времени суток
- 4. потребительская активность пассажиров зависит от авиакомпании, класса обслуживания (С/У), погоды в SVO и городе прилета (является ли курортом, расположен севернее/южнее/восточнее Москвы) и ряда других параметров рейса
- 5. У каждой торговой точки свой портрет целевой клиента
- 6. Пассажиры покупают преимущественно рядом со своими гейтами

## Если строить для каждой торговой точки отдельную модель, то можно составить портрет покупателя

К примеру, у точки №5 выручка зависит от времени суток, №20 зависит от направления вылетов, №21 от дня недели и сбоев в расписании, №9 от количества пассажиров, вылетающих из гейтов 1 этажа термина В). Все ~30 картинок доступны в репозитории на github.









# Прогноз влияния на выручку

Разработана математическая модель,  
предсказывающая выручку

Средняя точность прогнозов: 90%

По данным за 05.2022:

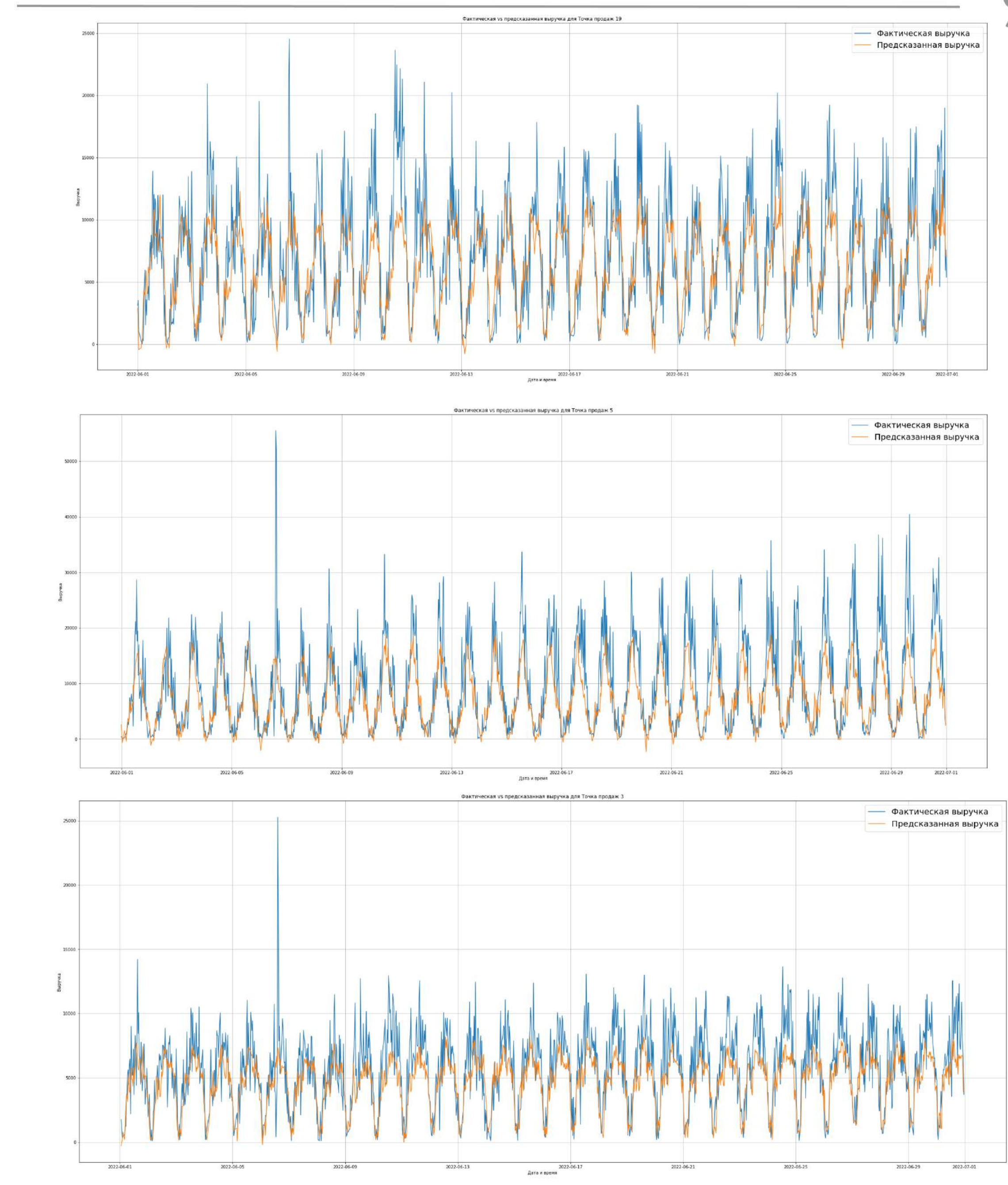
Средняя ошибка (MAE): 2092 у.е.

По данным за 06.2022:

Минимальная ошибка: 0.0 у.е

Средняя ошибка (MAE): 2786 у.е.

Максимальная ошибка: 94383 у.е.

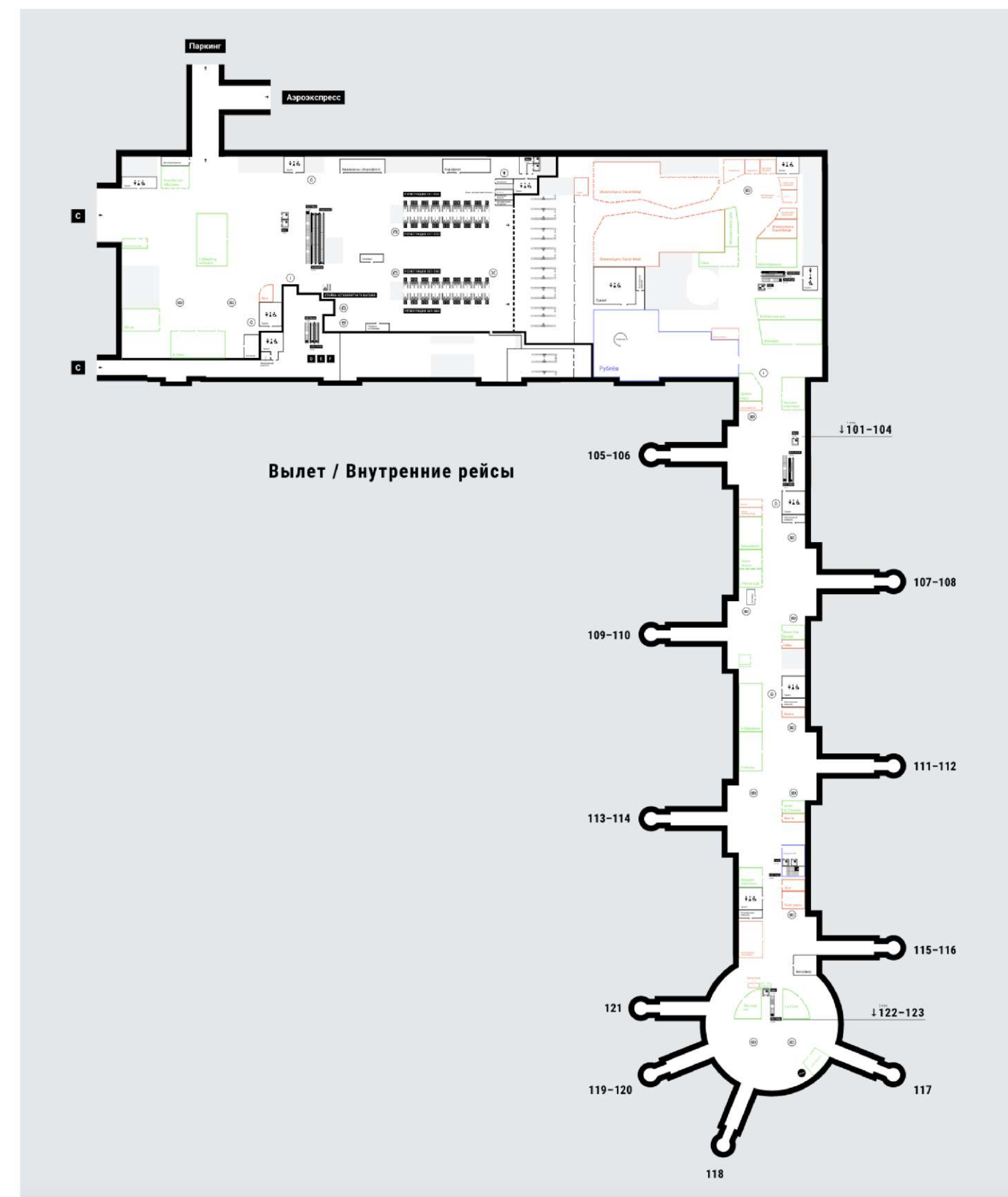




## Возможные управленческие решения

На основе полученных результатов предлагаю следующие решения:

1. Составить портрет покупателей каждой торговой точки
2. Моделировать и управлять потоками людей. У каждой точки свой “профиль”. Тогда можно назначать гейты на рейсы таким образом, чтобы рейс соответствовал торговой точке, находящимся рядом с гейтом. К примеру, бизнес-зал может быть рядом с рейсом Москва - Мюнхен, а магазин сувениров рядом с Москва-Анталья. В таком случае можно максимизировать выручку точки, как и SVO в целом.
3. Улучшить мобильность пассажиров между терминалами и внутри терминалов, чтобы пассажиры могли быстро достигать удаленной торговой точки и совершать в них покупки.





# Итоги





## Итоги

1. Проведен аналитический обзор предметной области, проанализированы предоставленные данные
2. Разработаны модели машинного обучения для прогнозирования выручки торговых точек
3. Определены ключевые признаки, которые влияют на выручка каждой точки
4. Разработанное решение позволяет в высокой точностью спрогнозировать выручку в будущем
5. Разработанное решение позволяет смоделировать выручку по более 40 входным параметрам (напр., revenue\_06.csv). Решение можно расширять.

## Контактные данные

Дробященко Игорь

Middle Data Scientist, S7 Airlines

e-mail: igordr1999@yandex.ru

telegram: @IgorCat1999



Исследование и описание данных:  
<https://github.com/Igordr1999/SVO>