

06	Об отчете				
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			

01.07.2022

## Описание концепции и подхода к работе с данными

Задача: спрогнозировать выручку торговых точек в Шереметьево

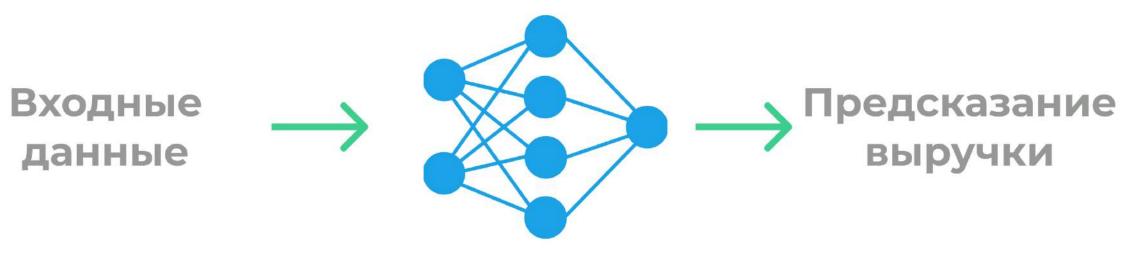
#### Концепция:

- 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

**Тренировочные Тестовые данные** 

Train = 1 месяц Test = 1 месяц

01.05.2022 01.06.2022

#### Пример данных для обучения ML модели

№ точки	Дата и время	Всего	AFL	SDM	C class	Y class	B term		Доход
14 10 1101	дата и время	pax	pax	pax	pax	рах	pax	***	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

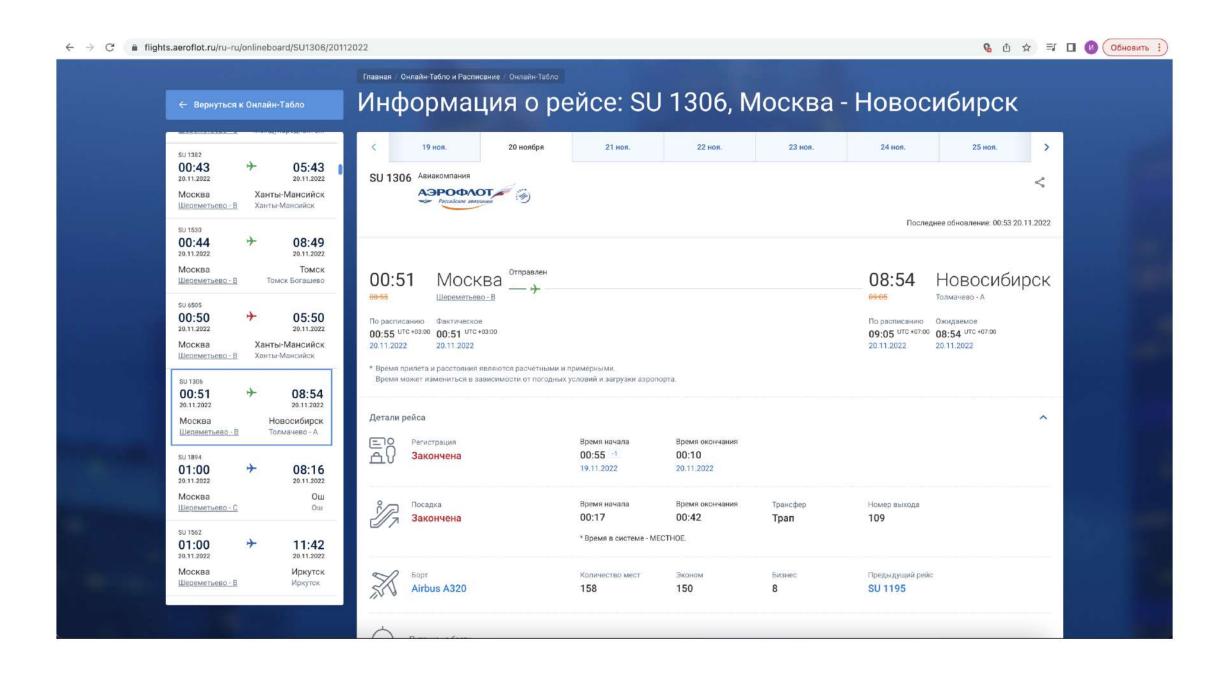
#### Используемые данные:

- 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 ісао и іата. С их помощью можно вычислить длительность полета, направление (юг или восток), страну прибытия
- 5. Список самолетов и их компановка (количество мест в бизнес и эконом классе)
- 6. Выручка торговых точек. Целевые данные.

#### Пример данных для обучения ML модели

№ точки	Дата и время	Всего	AFL	SDM	C class	Y class	B term		Доход
14- ТОЧКИ	дата и время	pax	pax	pax	pax	pax	pax	•••	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



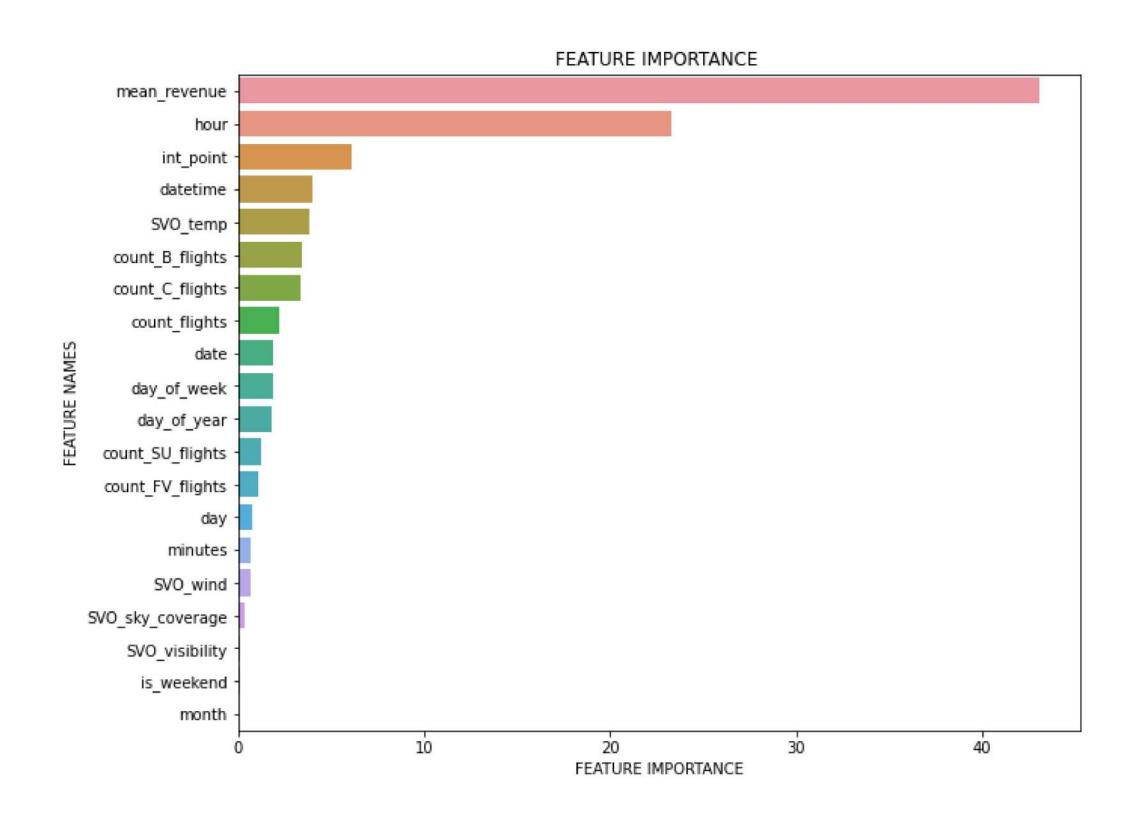
#### Описание гипотез

Ключевые гипотезы, которые "сработали":

- 1. Прямая зависимость между количеством пассажиров в чистой зоне в выручкой 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/Y), 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 выручкой 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/Y), 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

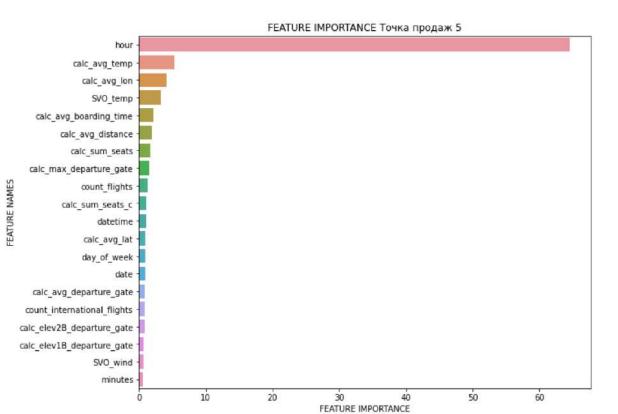
# Если строить для каждой торговой точки отдельную модель, то можно составить портрет покупателя

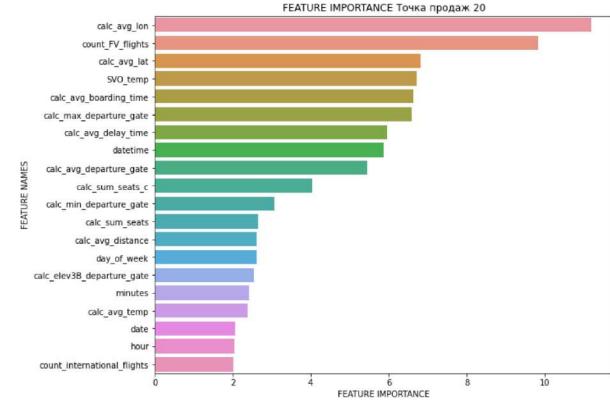
К примеру, у точки №5 выручка зависит от времени суток,

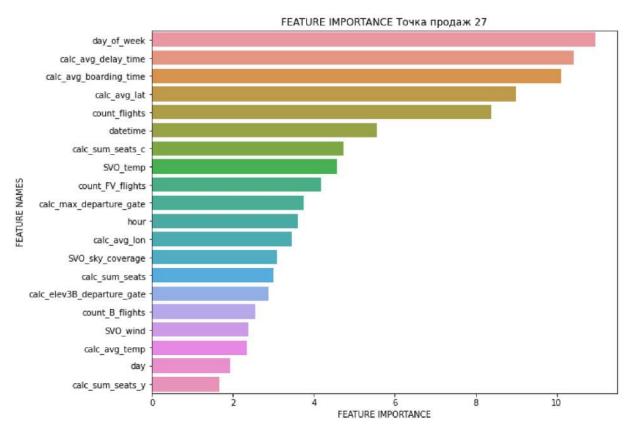
№20 зависит от направления вылетов,

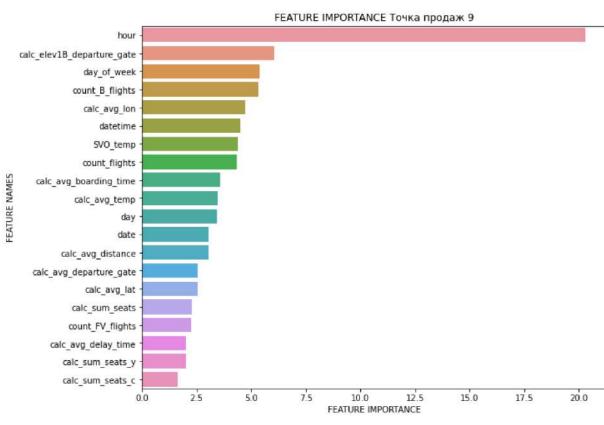
№21 от дня недели и сбоев в расписании,

№9 от количества пассажиров, вылетающих из гейтов 1 этажа термина В). Все ~30 картинок доступны в репозитории на github.









### Описание параметров для ML модели

1	Α	В	С	D	Е	F	G	Н	I	J
1	Параметр	Описание	Основной источник	Коммента	арий					
	date	Дата	Датасет Шереметьево							
	timeThirty	Дата и время	Датасет Шереметьево							
	revenue	Выручка	Датасет Шереметьево	Целевой	показател	ь				
	day_of_week	День недели	Получен из date	Выручка	в конце не	дели може	ет быть бол	тьше		
	day_of_year	Номер дня в году	Получен из date	Летом вы	іручка бол	ьше, чем в	есной			
	day	Номер дня в месяце	Получен из date							
	month	Номер месяца в году	Получен из timeThirty							
	hour	Количество часов	Получен из timeThirty							
	minutes	Количество минут	Получен из timeThirty							
	mean_revenue	Средняя выручка	Датасет Шереметьево	Средняя	выручка ка	аждой точн	ки (уровень	продаж)		
	is_weekend	Является ли выходным днем	Получен из date							
	timeHour	Количество часов	Получен из timeThirty							
	datetime	Дата и время	Получен из timeThirty							
,	SVO_temp	Температура воздуха в Москве	METAR	Возможн	а зависим	ость выруч	ки от пого,	ды в SVO		
	SVO_wind	Скорость ветра в Москве	METAR	Возможна зависимость выручки от погоды в SVO						
	SVO_visibility	Видимость в Москве	METAR	Возможна зависимость выручки от погоды в SVO						
	SVO_sky_coverage	Облачность в Москве	METAR	Возможн	а зависим	ость выруч	ки от пого,	ды в SVO		
	int_point	Номер точки продажи	Датасет Шереметьево							
	count_flights	Количество вылетающих рейсов в течение 3 часов	Онлайн табло							
	count_B_flights	Количество вылетающих рейсов из терминала В в течение 3 часов	Онлайн табло							
	count_C_flights	Количество вылетающих рейсов из терминала С в течение 3 часов	Онлайн табло							
	count_SU_flights	Количество вылетающих рейсов Аэрофтота в течение 3 часов	Онлайн табло							
	count_FV_flights	Количество вылетающих рейсов России в течение 3 часов	Онлайн табло							
	count_domestic_flights	Количество внутренних вылетающих рейсов в течение 3 часов	Онлайн табло							
	count_international_flights	s Количество международных вылетающих рейсов в течение 3 часов	Онлайн табло							
	calc_sum_seats	Количество вылетающих пассажиров в течение 3 часов	Онлайн табло							
	calc_sum_seats_c	Количество вылетающих пассажиров бизнес класса в течение 3 часов	Онлайн табло + данные компановки самолетов							
	calc_sum_seats_w	Количество вылетающих пассажиров комфорт класса в течение 3 часов	Онлайн табло + данные компановки самолетов							
	calc_sum_seats_y	Количество вылетающих пассажиров эконом класса в течение 3 часов	Онлайн табло + данные компановки самолетов							
	calc_avg_distance	Средняя дальность полета вылетающих рейсов в течение 3 часов	Данные аэропортов	Возможн	а зависим	ость выруч	ки от прод	олжительн	ость поле	га
	calc_avg_temp	Средняя температура в аэропортах прибытия для вылет. рейсов в течение 3 часов	METAR							
	calc_avg_lat	Средняя широта аэропортов прибытия для вылет. рейсов в течение 3 часов	Онлайн табло + данные аэропортов	Помогает	сформир	овать вект	ор вылетов	в (на юг, на	восток и т	д)
	calc_avg_lon	Средняя долгота аэропортов прибытия для вылет. рейсов в течение 3 часов	Онлайн табло + данные аэропортов	Помогает	сформир	овать вект	ор вылетов	в (на юг, на	восток и т	д)
	calc_avg_boarding_time	Средняя время посадки для вылет. рейсов в течение 3 часов	Онлайн табло							
	calc_avg_delay_time	Средняя задержка рейсов для вылет. рейсов в течение 3 часов	Онлайн табло	При заде	ржке рейс	ов увеличи	ивается вы	ручка в точ	ка продаж	
	calc_avg_departure_gate		Онлайн табло							
		Минимальный номер гейта для вылет. рейсов в течение 3 часов	Онлайн табло							
		Максимальный номер гейта для вылет. рейсов в течение 3 часов	Онлайн табло							
		а Количество вылетающих рейсов из гейтов 1 этажа термина С в течение 3 часов	Онлайн табло	The second secon				м от гейта ,		
		а Количество вылетающих рейсов из гейтов 2 этажа термина С в течение 3 часов	Онлайн табло					м от гейта ,		
	The state of the s	а Количество вылетающих рейсов из гейтов 3 этажа термина С в течение 3 часов	Онлайн табло	100				м от гейта ,		
		а Количество вылетающих рейсов из гейтов 1 этажа термина В в течение 3 часов	Онлайн табло					м от гейта ,		
		а Количество вылетающих рейсов из гейтов 2 этажа термина В в течение 3 часов	Онлайн табло					м от гейта ,		
	calc_elev3B_departure_ga	а Количество вылетающих рейсов из гейтов 3 этажа термина В в течение 3 часов	Онлайн табло	Большая	корреляці	ия между р	асстояние	м от гейта ,	до точки п	родаж

#### Прогноз влияния на выручку

Разработана математическая модель, предсказывающая выручку

Средняя точность прогнозов: 90%

По данным за 05.2022:

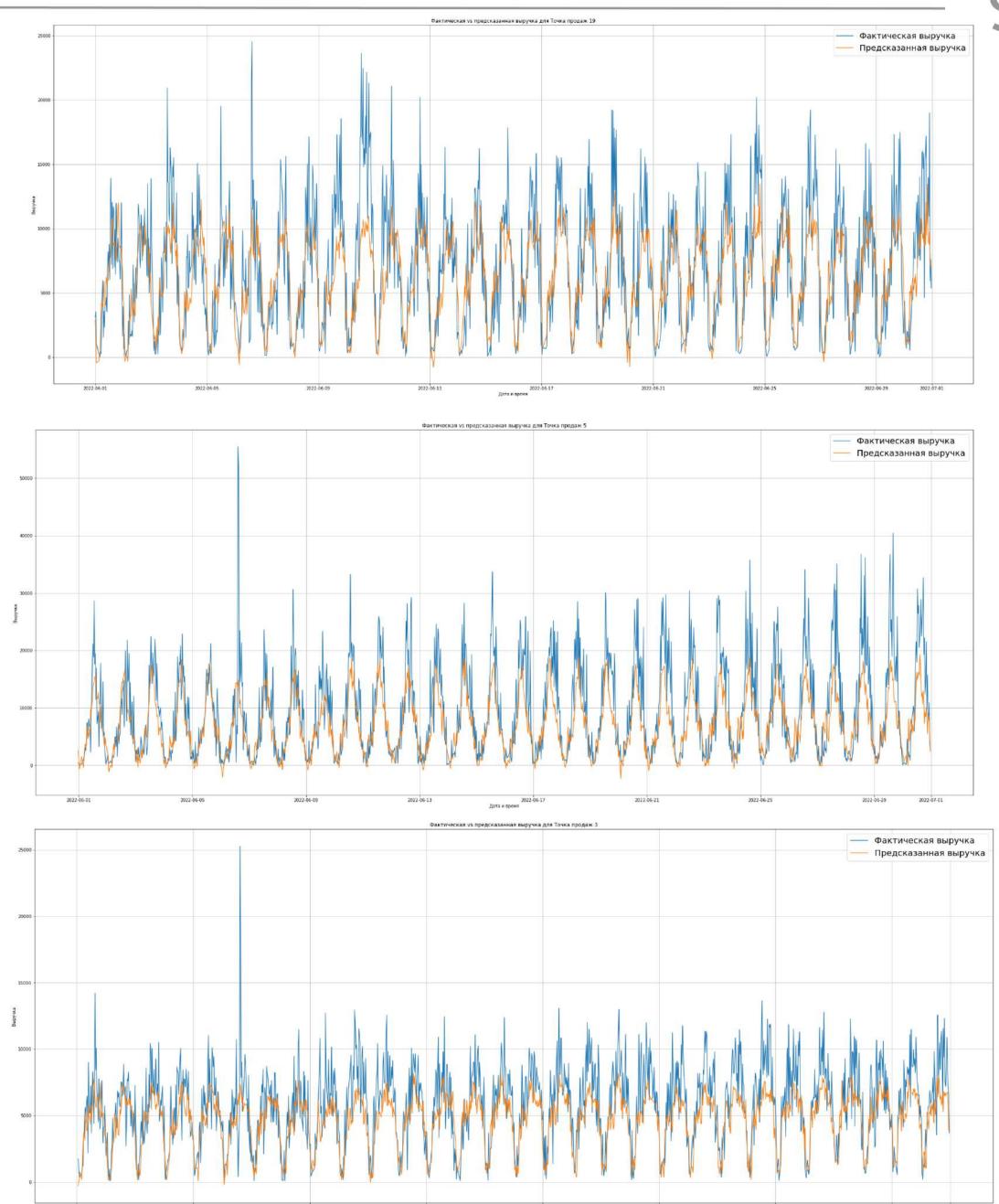
Средняя ошибка (МАЕ): 2092 у.е.

По данным за 06.2022:

Минимальная ошибка: 0.0 у.е

Средняя ошибка (МАЕ): 2786 у.е.

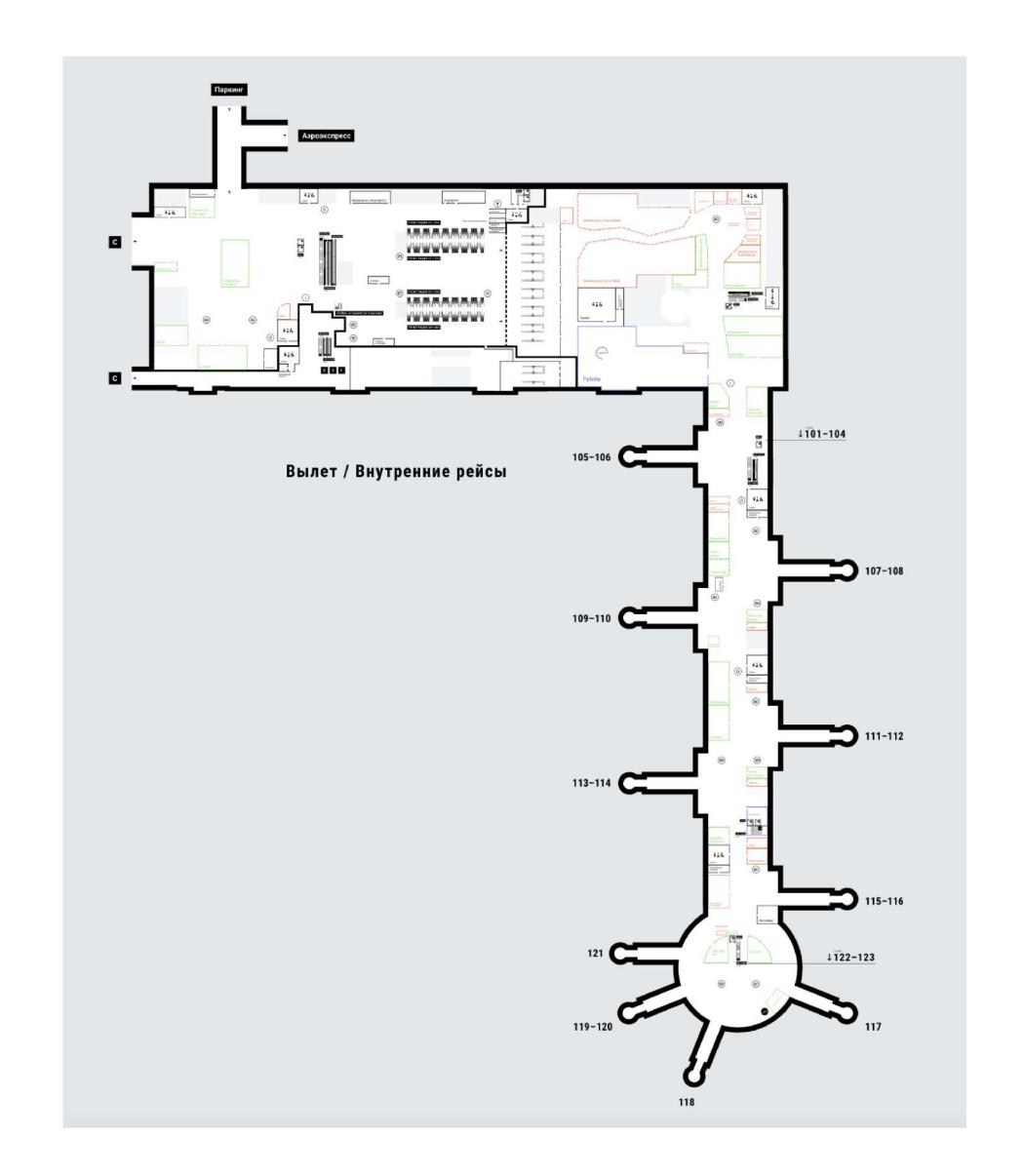
Максимальная ошибка: 94383 у.е.



#### 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 соотвествовал 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: @lgorCat1999



Исследование и описание данных: https://github.com/lgordr1999/SVO