

ЗАДАЧИ

- 1. На основании данных первого месяца из датасета (файлы 05.2022_Выручка и 05.2022_Пассажиропоток), построить гипотезы по работе с данными и сегментации пассажиропотока в зависимости от различных факторов, таких как география рейсов, авиакомпания, время ожидания пассажира в «чистой зоне».
- 2. Определить сегменты, авиакомпании и направления, с которыми связана максимальная потребительская активность.
- 3. Оценить влияние иных факторов, таких как задержки рейсов, погодные условия и другие факторы.
- 4. Построить прогноз выручки на второй месяц из датасета на основании данных пассажиропотока (файл 06.2022_Пассажиропоток).
- 5. Выработать рекомендации по увеличению выручки в зависимости от доступных факторов и сегментации пассажиропотока

ВХОДНЫЕ ДАННЫЕ

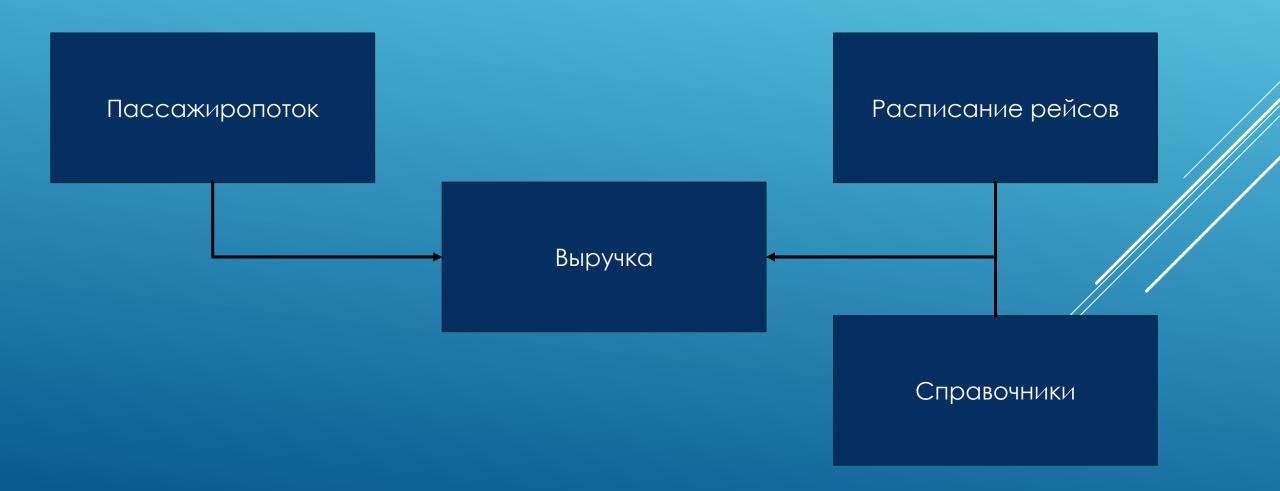
- Выручка арендаторов с привязкой к временным интервалам.
- Информация о плановой и фактической дате вылета, отметка о случае отмены рейса.
- ▶ Обезличенная информация о находящихся в «чистой зоне» пассажирах.



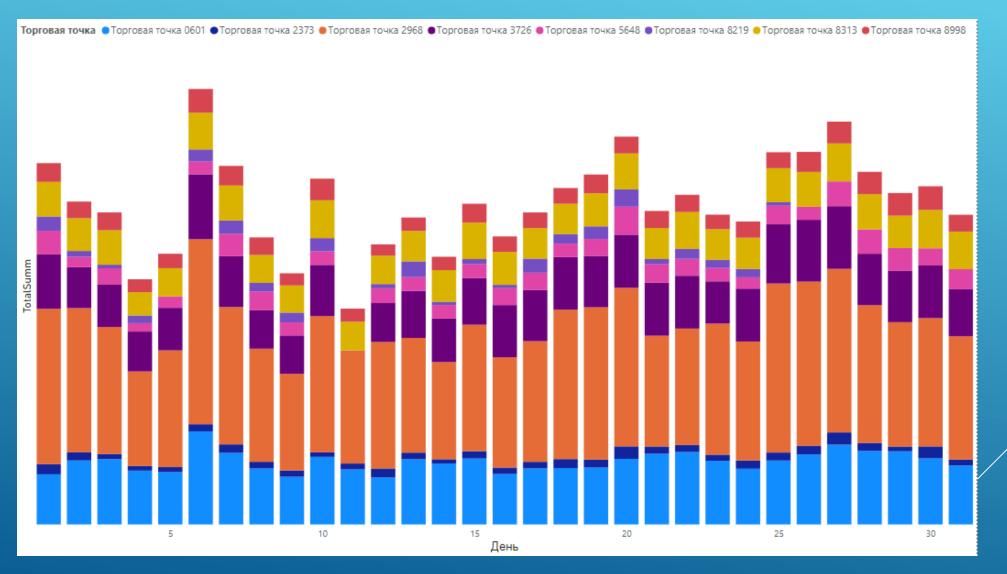


1. Гипотезы по работе с данными и сегментации пассажиропотока.

- 1. Гипотеза по работе с данными заключается в том, что нам необходимо проанализировать файл с выручкой, определить точки с максимальной выручкой, далее сопоставить с полученными данными данные о пассажиропотоке, на основе этого выявить тип пассажира, который приносит эту выручку, его предпочтения, на основе информации о рейсах и из справочников получить полную картину для принятия управленческих решений и построить на основе данных о пассажиропотоке следующего месяца прогноз по выручке.
- 2. Выявить какие данные оказывают наибольшее влияние на выручку.

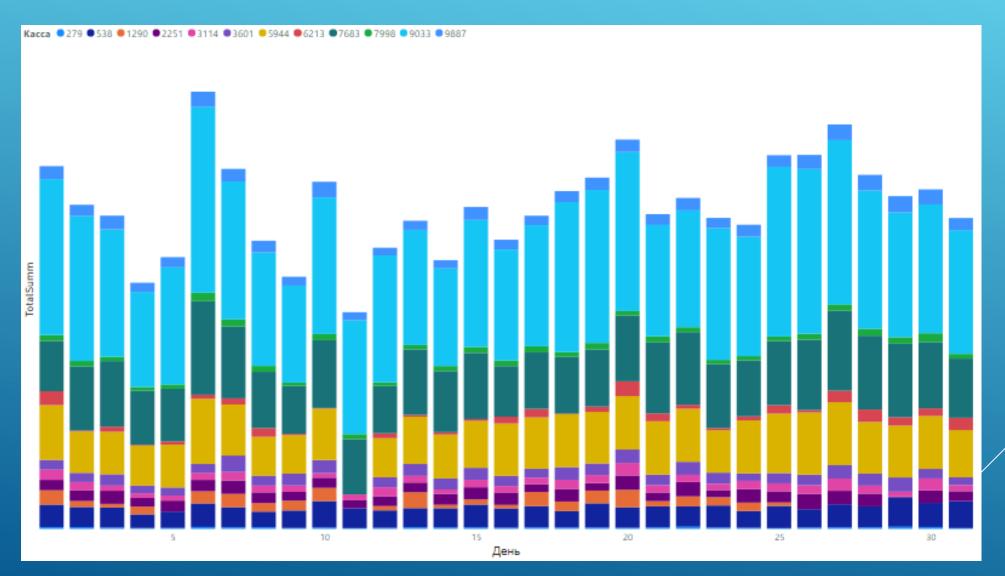


Сегментация выручки торговых точек за май



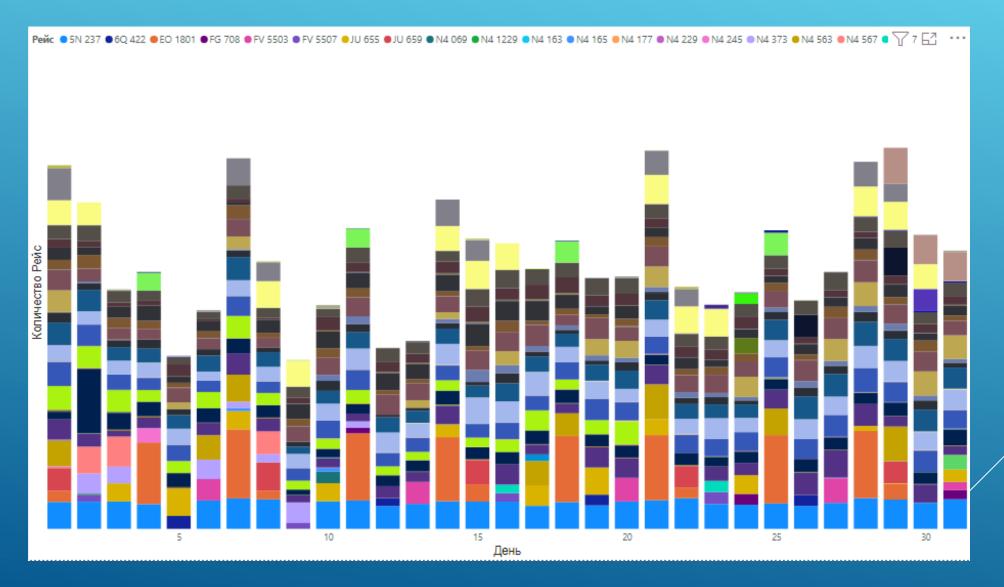
На данном графике представлена доля выручки каждой торговой точки в общей выручке по дням за май месяц, так же мы увидеть лидеров по выручке.

Сегментация выручки касс торговых точек за май



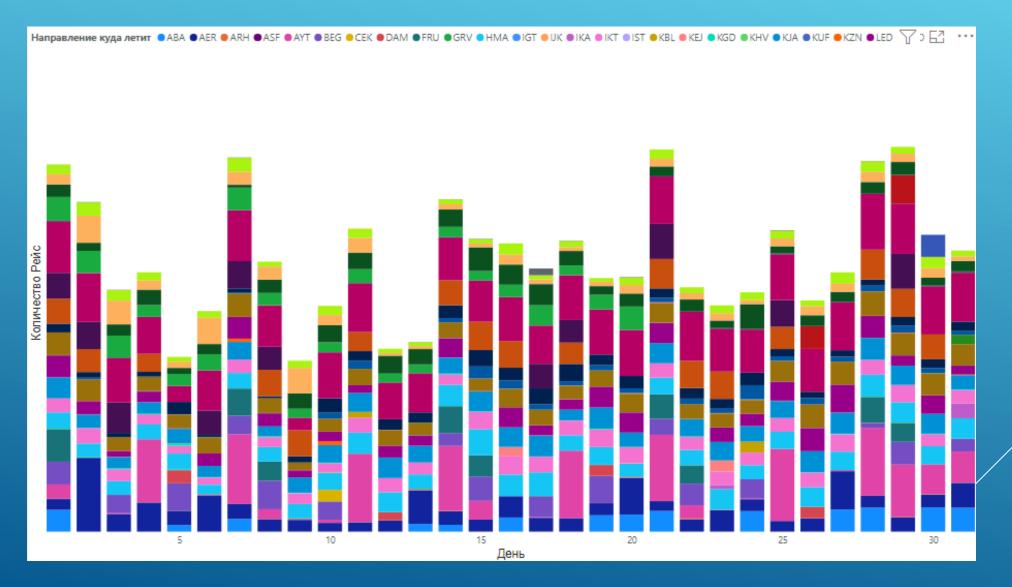
На данном графике представлена доля выручки каждой кассы торговых точек в общей выручке по дням за май месяц, так же мы увидеть лидеров по выручке.

Сегментация рейсов в пассажиропотоке за май



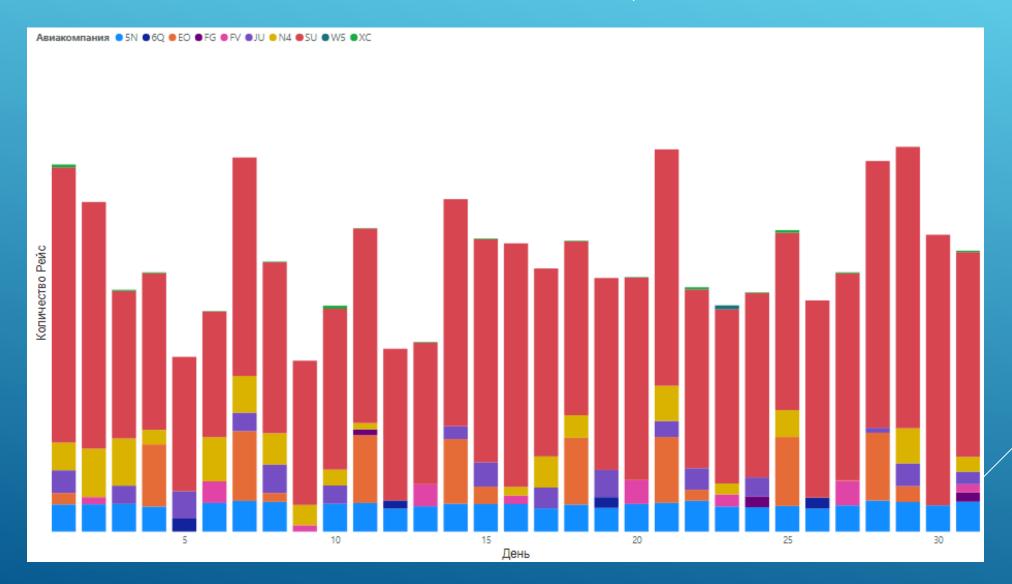
На данном графике представлена сегментация рейсов в общем пассажиропотоке по дням за май месяц, так же мы увидеть каких рейсов было больше всего конкретный день.

Сегментация направлений рейсов в пассажиропотоке за май



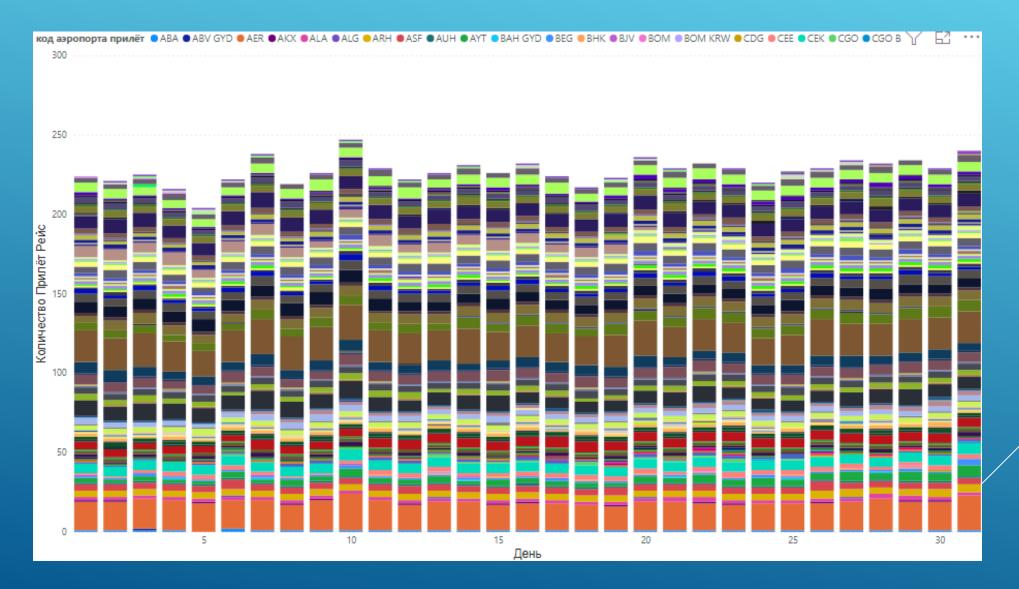
На данном графике представлена информация о направлениях рейсов по дням за май месяц, так же мы увидеть основные направления рейсов.

Сегментация по авиакомпаниям в пассажиропотоке за май



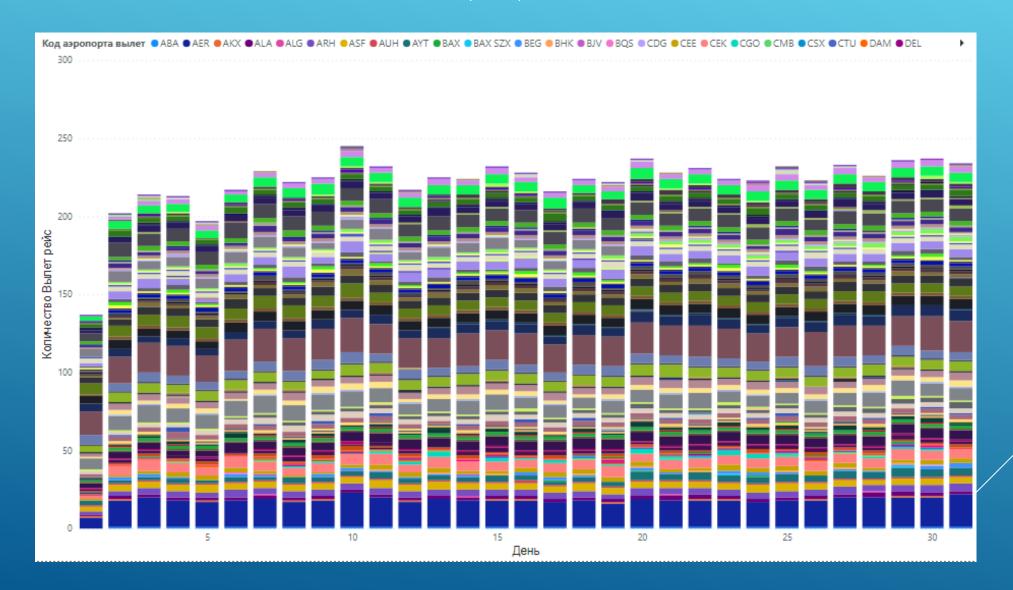
На данном графике представлена информация об авиакомпаниях по дням за май месяц, так же мы увидеть авиакомпании выполняющие наибольшее количества

Сегментация аэропортов прилетов за май



На данном графике представлена информация об аэропортах с наибольшим количеством прилетов по дням за май месяц.

Сегментация аэропортов вылетов за май



На данном графике представлена информация об аэропортах с наибольшим количеством вылетов по дням за май месяц.



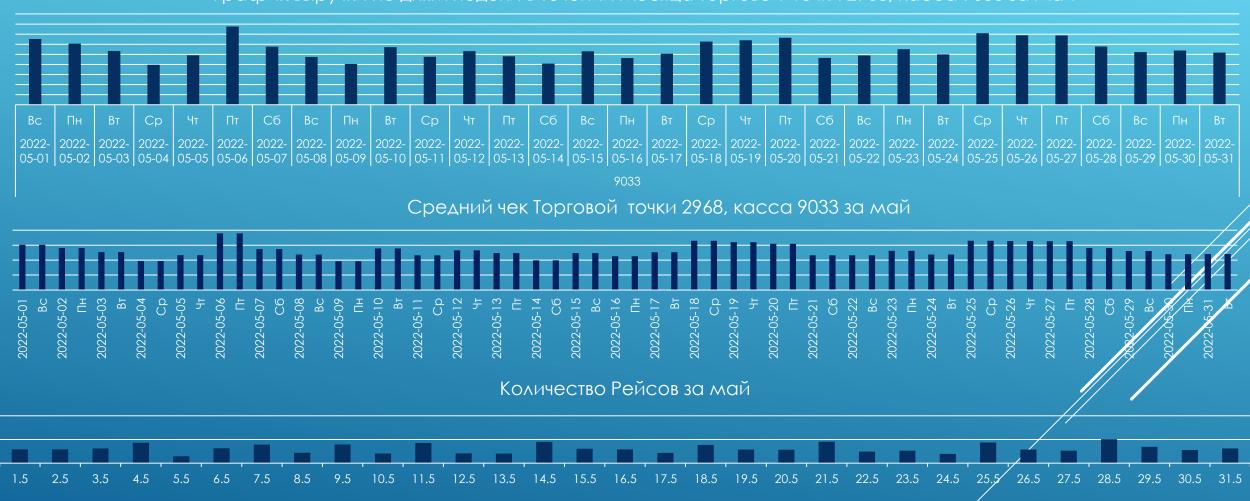
Выводы: Считаю на эти точки нужно обратить особое внимание, для переноса их практики продажи на другие точки, с целью увеличения выручки отстающих точек, так же интересен график среднего чека, мы видим высокий средний чек в кассах 6213, 1290.

Таблица выручки, кол-во операций, и суммы среднего чека торговых точек за май

Торговые точки	Сумма по полю TotalSumm	Кол-во операций	Средний чек
Общепит	7 756 761 842,40	9 214,00	841 845,22
Торговая точка 0601	1 635 165 200,80	1 378,00	1 186 622,06
7683	1 635 165 200,80	1 378,00	1 186 622,06
Торговая точка 2373	192 717 080,00	968,00	199 087,89
3114	192 717 080,00	968,00	199 087,89
Торговая точка 2968	3 408 585 641,60	1 285,00	2 652 595,83
9033	3 408 585 641,60	1 285,00	2 652 595,83
Торговая точка 3726	1 248 330 640,00	1 322,00	944 274,31
5944	1 248 330 640,00	1 322,00	944 274,31
Торговая точка 5648	419 581 920,00	1 358,00	308 970,49
3601	274 386 160,00	1 108,00	247 640,94
6213	145 195 760,00	250,00	580 783,04
Торговая точка 8313	852 381 360,00	2 903,00	293 620,86
0279	37 245 600,00	534,00	69 748,31
0538	536 254 640,00	1 371,00	391 141,24
2251	278 881 120,00	998,00	279 440,00
Торговля	624 252 080,00	2 626,00	237 719,76
Торговая точка 8219	174 944 000,00	152,00	1 150 947,37
1290	174 944 000,00	152,00	1 150 947,37
Торговая точка 8998	449 308 080,00	2 474,00	181 612,00
7998	141 844 880,00	1 186,00	119 599,39
9887	307 463 200,00	1 288,00	238 713,66
Общий итог	8 381 013 922,40	11 840,00	707 855,91



График выручки по дням недели в течении месяца Торговой точки 2968, касса 9033 за май



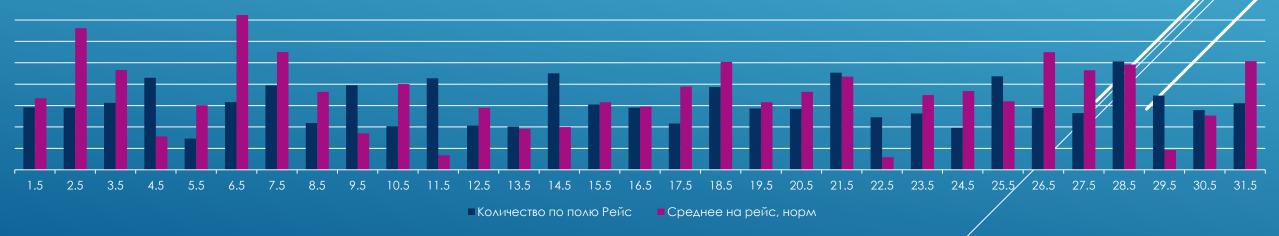
Из данных графиков мы видим, выручка и средний чек в этой торговой точки коррелируют, нет выбросов, здесь мы можем поднять прибыть увеличив средний чек, так же мы видим, что кол-во рейсов в конкретный день и выручка этой точки не связанны, значит нужно смотреть именно структуру рейсов.

График выручки в течении месяца и количества рейсов в конкретное время за май



Из данного графика мы видим, что нет связи с количество рейсов и выручкой по времени, значит нужно смотреть именно структуру рейсов в конкретный день, для понимания типа пассажира

График средней выручки в течении месяца на количество рейсов в конкретное время за май



Далее рассмотрим подробней лидеров по средней выручки за рейс.

Таблица лидеров выручки, кол-во операций, и суммы среднего чека торговых точек за май

		10110110101	0000000	511 5 GI 77 1 GI7 1		
Выход на рейс (временем	Количество по полю Рейс	Выручка	Рейс	Авиакомпания	Направление куда летит	Разница входа в чистую зону и выхода на рейс
06.05.22 18	3:00 1	14 453 104,80	SU 1700	SU	VVO	17:49:56
02.05.22 14	:00 1	13 236 300,00	SU 1078	SU	IGT	23:03:06
07.05.22 12	1:30 1	11 007 172,80	SU 1364	SU	STW	16:28:44
26.05.22 14	:00 1	10 984 609,60	SU 1506	SU	TJM	14:15:47
07.05.22 7	':00 1	10 857 166,40	SU 1320	SU	MMK	7:17:11
31.05.22 16	:00 1	10 148 720,00	SU 1700	SU	VVO	16:44:00
18.05.22 8	3:30 1	10 079 436,00	SU 1200	SU	PEE	9:00:58
28.05.22 14	:00 1	9 831 920,00	SU 1422	SU	CEK	17:18:07
26.05.22 9	:00 1	9 831 600,00	SU 1486	SU	KJA	10:21:32
28.05.22 11	:30 1	9 514 527,20	SU 1500	SU	TJM	14:30:06
03.05.22 16	:00 1	9 328 482,40	SU 1700	SU	VVO	17:30:53
27.05.22 9	1:00	9 292 824,00	SU 1366	SU	STW	12:24:31
06.05.22 11	:30 1	8 748 896,00	SU 016	SU	LED	18:23:07
07.05.22 9	:00 1	8 704 544,00	SU 1366	SU	STW	17:26:14
21.05.228	:30 1	8 697 713,60	SU 1830	SU	MSQ	11:26:18
28.05.22 12	2:00 1	8 277 074,40	SU 1642	SU	ASF	13:39:35
10.05.22 17	':00 1	8 011 433,60	SU 1730	SU	PKC	17:29:03
17.05.22 12	1:30	7 769 296,00	SU 1364	SU	STW	12:36:30
03.05.22 18	3:00 1	7 762 720,00	SU 1730	SU	PKC	22:04:50
24.05.22 17	':00 1	7 347 832,00	SU 1730	SU	PKC	17:25:13
08.05.22 9	:30 1	7 275 200,00	SU 1190	SU	KZN	10:53:42
20.05.22 17	':00 1	7 271 707,20	SU 1730	SU	PKC	20:07:33
23.05.22 7	:00 1	6 985 600,00	SU 1390	SU	SCW	7:52:07
10.05.22 8	:30 1	6 906 640,80	SU 1200	SU	PEE	9:06:53
02.05.22 7	:30 1	6 768 452,80	SU 1886	SU	FRU	8:51:07
03.05.22 9	:00 1	6 703 002,40	SU 1366	SU	STW	14:19:15
01.05.22 9	:00 1	6 679 603,20	SU 1366	SU	STW	9:09:22
25.05.22 6	:30 1	6 396 320,00	SU 1000	SU	KGD	8:08:45
01.05.22 10	:00 1	6 354 506,40	SU 1002	SU	KGD	16:21:15

Выводы: По данным анализа мы видим, что выручка торговых точек коррелирует с временем нахождения в чистой зоне пассажиров

Работа с данными:

- 1. Предоставленные данные были объединены в один датасет с учетом выручки торговых точек и касс по времени.
- 2. Для данных из датасета «пассажиропоток» рассчитана разница между выходом пассажиров в чистую зону и выходом на рейс. Объединение с датасетом «Выручка» по времени выхода на рейс по времени округленным вниз до получасового интервала для сопоставления с временем выручки касс.
- 3. Для данных из датасета «Расписание рейсов» получена разница между временем приземления и временем вылета. Объединение с датасетом «Выручка» по времени вылета округленным до получасового интервала для сопоставления с временем выручки касс.

Дополнение:

При объединении датасетов были обнаружены пропуски по времени в значениях из датасетов пассажиропотока и расписания рейсов, отсутствие рейсов во времени получения выручки, выручка есть а пассажиров нет в терминале.

С	D	E	F	G	Н	1	J	K	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
Торговая точка	Касса	orgtype	termin al	tzone	TotalSumm	День недели	TimeThirty2		Кол-во рейсов вылет	Средне е время ожидан ия	Среднее время ожидания в часах	Суммарн ое время ожидания в часах	Суммар ное время ожидан ия	Макс время ожидан ия	Кол-во рейсов вылет по расспис анию	Средне е время ожидан ия между прилет и	Средне е время ожидан ия в часах между	Суммар ное время ожидан ия в между	Суммар ное время ожидан ия в часах	Макс время ожидан ия между прилет и
▼	~	~	~		~	~	~	7 🔻	~	~	~	~	~	~	~	вылет	вылет	вылет	прилет и	вылет
Горговая точка 0601	7683	Общепит	В	Чистая зона ВВЛ-вылет	755 448,0	Вс	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 8313	2251	Общепит	В	Чистая зона ВВЛ-вылет	44 560,0	Вс	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 8313	0538	Общепит	В	Чистая зона ВВЛ-вылет	242 160,0	Bc	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 2968	9033	Общепит	В	Чистая зона ВВЛ-вылет	712 000,0	Bc	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 3726	5944	Общепит	В	Чистая зона ВВЛ-вылет	44 000,0	Вс	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 5648	3601	Общепит	В	Чистая зона ВВЛ-прилет	387 920,0	Вс	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 2373	3114	Общепит	В	Общедоступная зона-вылет	38 800,0	Bc	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Горговая точка 8998	7998	Торговля	В	Чистая зона ВВЛ-вылет	59 600,0	Bc	01.05.22 0:00	2	583	0,07809	1:52:27	12:36:01	45,525	0,28947	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д

3. Влияние иных факторов, таких как задержки рейсов, погодные условия и другие факторы.

Для анализа и предсказания выручки добавлены данные о погоде с разбивкой на получасовые интервалы и объединены с общим датасетом, выполнены дополнительные расчеты по времени нахождения пассажиров в аэропорту.

Исследуя такой параметр как, важность данных для предсказания модели, можно утверждать, что данные о погоде оказывают влияние на выручку касс и торговых точек.

Основные обозначения:

Т, температура воздуха (градусы Цельсия) на высоте 2 метра над поверхностью земли

РО, атмосферное давление на уровне станции (миллиметры ртутного столба)

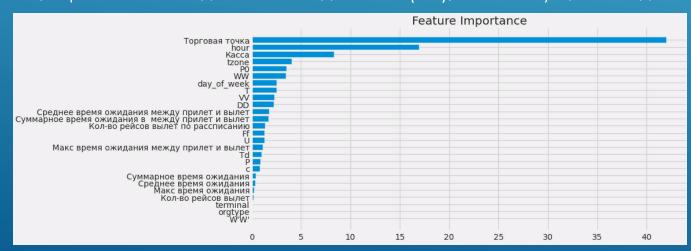
Р, атмосферное давление, приведенное к среднему уровню моря (миллиметры ртутного столба)

U, относительная влажность (%) на высоте 2 метра над поверхностью земли

DD, направление ветра (румбы) на высоте 10-12 метров над земной поверхностью, осредненное за 10-минутный период, непосредственно предшествовавший сроку наблюдения

FF, скорость ветра на высоте 10-12 метров над земной поверхностью, осредненная за 10-минутный период, непосредственно предшествовавший сроку наблюдения (метры в секунду)

VV, горизонтальная дальность видимости (км), WW – текущая погода.



4. Прогноз выручки на второй месяц из датасета на основании данных пассажиропотока (файл 06.2022_Пассажиропоток).

Прогноз выручки на 06.2022 рассчитан по предоставленным данным, с использованием двух разных моделей прогноза регрессии: «CatBoostRegressor» и «lightgbm»

Для оценки точности моделей регрессии выбрана метрика: Средняя абсолютная ошибка (МАЕ).

МАЕ ошибка рассчитывается как среднее абсолютных разностей между целевыми значениями и прогнозами.

Для модели «CatBoostRegressor» Средняя абсолютная ошибка равна:

MAE: 353270.0749009358

Для модели «lightgbm» Средняя абсолютная ошибка равна:

MAE: 364537.8215109858

Рекомендация: Для прогнозирования выручки с учетом предоставленных и дополнительных данных рекомендуется использовать модель «CatBoostRegressor», так как её точность её предсказания выше других моделей, для более точного прогноза необходимо заполнить пропуски реальными данными.

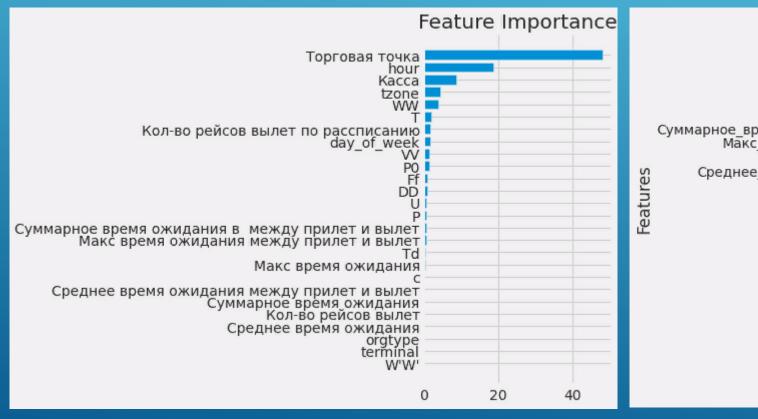
$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

5. Рекомендации по увеличению выручки в зависимости от доступных факторов и сегментации пассажиропотока

Исследуя такой параметр как, важность данных для предсказания модели, можно увидеть какие параметры из датасета оказывают влияние на выручку касс и торговых точек.

График влияния данных на прогноз выручки модели «CatBoostRegresson»

График влияния данных на прогноз выручки модели «lightgbm»





СПАСИБО ЗА ВНИМАНИЕ!