

Task 1

برای این قسمت از query زیر استفاده شد:

```
SELECT
  s.store_id,
  s.store_name,
  COUNT(o.order_id) AS total_orders,
  COUNT(CASE WHEN o.order_status='FINISHED' THEN 1 END) AS finished_orders,
  (COUNT(CASE WHEN o.order_status='FINISHED' THEN 1 END)*100/COUNT(o.order_id)) AS percentage_net_orders,
  AVG(o.order_amount) AS average_order_amount,
  SUM(p.payment_amount) AS total_sales,
  AVG(d.delivery_distance_meters) AS average_delivery_distance,
  (COUNT(CASE WHEN dr.driver_modal = 'MOTOBOY' THEN 1 END) * 100 / COUNT(CASE WHEN dr.driver_modal IS NOT null THEN 1
END)) AS percentage_delivered_by_motoboy,
  (100 - (COUNT(CASE WHEN dr.driver_modal = 'MOTOBOY' THEN 1 END) * 100 / COUNT(CASE WHEN dr.driver_modal IS NOT null
THEN 1 END))) AS percentage_delivered_by_biker,
FROM orders o
LEFT JOIN stores s ON s.store_id = o.store_id
LEFT JOIN payments p ON o.payment_order_id = p.payment_order_id
LEFT JOIN deliveries d ON o.delivery_order_id = d.delivery_order_id
LEFT JOIN drivers dr ON d.driver_id = dr.driver_id
GROUP BY s.store_id, s.store_name
ORDER BY s.store_id
```

1	SELECT
2	s.store_id,
3	s.store_name,
4	COUNT(o.order_id) AS total_orders,
5	COUNT(CASE WHEN o.order_status='FINISHED' THEN 1 END) AS finished_orders,
6	(COUNT(CASE WHEN o.order_status='FINISHED' THEN 1 END)*100/COUNT(o.order_id)) AS percentage_net_orders,
7	AVG(o.order_amount) AS average_order_amount,
8	SUM(p.payment_amount) AS total_sales,
9	AVG(d.delivery_distance_meters) AS average_delivery_distance,
10	(COUNT(CASE WHEN dr.driver_modal = 'MOTOBOY' THEN 1 END) * 100 / COUNT(CASE WHEN dr.driver_modal IS NOT null THEN 1 END)) AS percentage_delivered_by_motoboy,
11	(100 - (COUNT(CASE WHEN dr.driver_modal = 'MOTOBOY' THEN 1 END) * 100 / COUNT(CASE WHEN dr.driver_modal IS NOT null THEN 1 END))) AS percentage_delivered_by_biker,
12	FROM orders o
13	LEFT JOIN stores s ON s.store_id = o.store_id
14	LEFT JOIN payments p ON o.payment_order_id = p.payment_order_id
15	LEFT JOIN deliveries d ON o.delivery_order_id = d.delivery_order_id
16	LEFT JOIN drivers dr ON d.driver_id = dr.driver_id
17	GROUP BY s.store_id, s.store_name
18	ORDER BY s.store_id

store_id	store_name	total_orders	finished_orders	percentage_net_o...	average_order_a...	total_sales	average_delivery_...	percentage_delive...
3	Store1	13	0	0	47.96923076923076	200260.47999999929	2714.1638459651704	76
6	Store2	4100	4012	97	52.45763414633912	110617.27999999972	5491.655963250548	68
8	Store3	2744	2569	93	42.027015306122486	1043532.8600000188	1406.0354166666666	34
53	Store4	16924	16821	99	66.52300519971936	7146.680000000004	1247.6039603960396	35
54	Store5	101	96	95	77.53465346534647	156752.189999999508	1153.4084935897436	27
56	Store6	6659	6569	98	38.140339390298244	44.0392523364485...	8418.499999999999	1382.3804347826087
58	Store7	214	167	78	44.0392523364485...	111563.00000000048	5499.7543398211465	48
82	Store8	1949	1773	90	57.12404822986114	350.9999999999999...	1091.1111111111111	25
83	Store9	24	8	33	36.620833333333316	53.258966897051735	1677.120499632623	49
84	Store10	2749	1964	71	59.09818965517245	22436.49999999999...	1398.4193548387098	35
85	Store11	348	339	97	116.53409090909086	5303.729999999998	2993.888888888888...	48
88	Store12	66	64	96	46.021925675675675	12828.5	2863.752542372881	38
89	Store13	296	283	95	53.52365217391308	14219.139999999952	1764.8496240601503	45
90	Store14	460	418	90	50.72400000000001	58696.46000000092	1559.1025641025642	33
95	Store15	1170	1120	95	40.90820895522389	103323.19999999943	1846.6140776699028	45
96	Store16	67	0	0	82.12967741935424	151142.96999999987	1616.857095046854	38
99	Store17	1240	1221	98	60.51668025078241			
100	Store18	3190	3142	98				

تحلیل:

اگر بر حسب بیشترین فروش (total sales) مرتب کنیم.

	store_name	total_orders	finished_orders	percentage_net_o...	average_order_a...	total_sales	average_delivery_...	percentage_delive...	percentage_delive...
69	Store134	13337	13284	99	133.59046787133448	1876241.1200000274	2143.571944068561	89	11
70	Store153	10412	10323	99	136.53353246254485	1483513.4900000163	2446.907728246058	87	13
71	Store33	8849	8793	99	132.50570346932201	1230023.4100000062	2462.803233256351	82	18
72	Store126	7972	7898	99	140.51437782238133	1151186.2200000035	2273.586871189964	79	21
73	Store22	8709	8611	98	126.82911356068871	1139415.9500000023	1280.6654183733178	26	74
74	Store4	16924	16821	99	66.52300519971936	1043532.8600000188	1406.0354166666666	34	66
75	Store29	7793	7721	99	125.59874502759374	1016699.7200000002	2513.825841219769	68	32
76	Store37	7286	7132	97	123.5735286851545	921614.75999999974	1950.2432995417303	76	24
77	Store267	6826	6722	98	126.63219308526658	905777.14999999949	2632.4819564569043	81	19
78	Store76	6493	6458	99	125.59915755429357	856830.0400000074	1446.23686664594...	52	48
79	Store61	5544	5464	98	136.37666486291764	787400.7800000018	3114.2689352360044	99	1

میزان تمام شدن سفارش بالا 98 درصد بوده. میزان متوسط سفارش در آن ها حدوداً ۱۳۰ و تعداد سفارش ها حدوداً ۸۰۰۰ (۵۰۰۰ تا ۱۳۰۰۰) بوده است.

در صورتی که بر حسب میزان متوسط سفارش مرتب کنیم:

	store_name	total_orders	finished_orders	percentage_net_o...	avgerage_order_a...	total_sales	average_delivery_...	percentage_delive...	percentage_delive...
1	Store360	2	2	100	7499	7499	243	0	100
2	Store102	295	280	94	6967.2143728813535	265283.3799999999...	4686.36170212766	92	8
3	Store626	1	1	100	6399	6399	7049	100	0
4	Store864	48	46	95	5287.736666666667	246244.36	6201.9375	97	3
5	Store450	5	3	60	3399.4	17016.8	2708	100	0
6	Store302	2	2	100	3199	6398	250	50	50
7	Store362	58	58	100	2697.4191379310346	156450.31	3670.137931034483	78	22
8	Store561	23	23	100	2496.14347826087	57411.3	4393.130434782609	77	23

تعداد سفارش ها بسیار کمتر (کمتر از ۱۰۰) بوده و میزان فروش کلی از کم تا زیاد موجود است. اگر بر حسب متوسط average_delivery_distance مرتب کنیم:

	store_name	total_orders	finished_orders	percentage_net_o...	avgerage_order_a...	total_sales	average_delivery_...	percentage_delive...	percentage_delive...
32	Store82	1	0	0	25		3295201		
33	Store327	36	19	52	56.83333333333333...	1432.2900000000002	3166989.1764705884	90	10
34	Store573	552	550	99	88.204891304348	39200.40000000025	1941699.9799635701	90	10
35	Store508	2	1	50	179.9	179.91	1447909		
36	Store189	827	643	77	85.30519951632417	46198.99000000011	954407.4729559749	74	26
37	Store19	2	0	0	68.7		846739.5		
38	Store633	2	2	100	50	100	547893	100	0
39	Store548	32	30	93	32.7625	1052.2999999999997	495344.78571428574	88	12
40	Store756	1	0	0	250		340269		
41	Store96	664	584	87	100.61530120481963	46662.54000000025	334151.98267716536	62	38
42	Store591	5	2	40	52.179999999999999	49.9	272334.8	100	0
43	Store434	12	12	100	257.7291666666667	3154.25	223628.666666666...	81	19

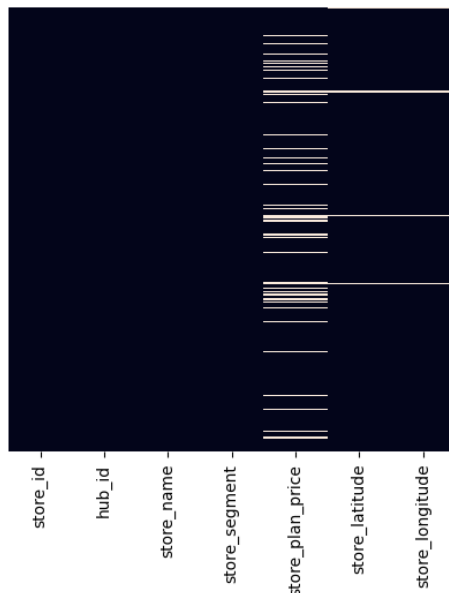
طبق انتظار درصد motor boy بیشتر از biker بوده است. درصد سفارشات کنسل شده بیشتر شده است.

Task 2

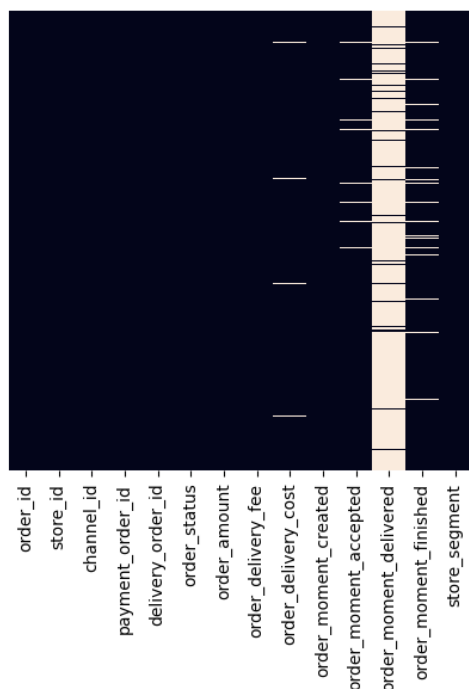
در ابتدا به بررسی store.csv پرداخته و

```
store_segment
GOOD 567
FOOD 384
Name: count, dtype: int64
```

در این فایل تعداد GOOD و FOOD را ملاحظه میکنید. برای بررسی null ها در این فایل:



بخش store_segment که شامل GOOD و FOOD است هیچ مقدار null ندارد. از left join استفاده شد تا orders را با store ادغام کنیم و برای بررسی مقادیر null نیز شکل زیر را داریم:

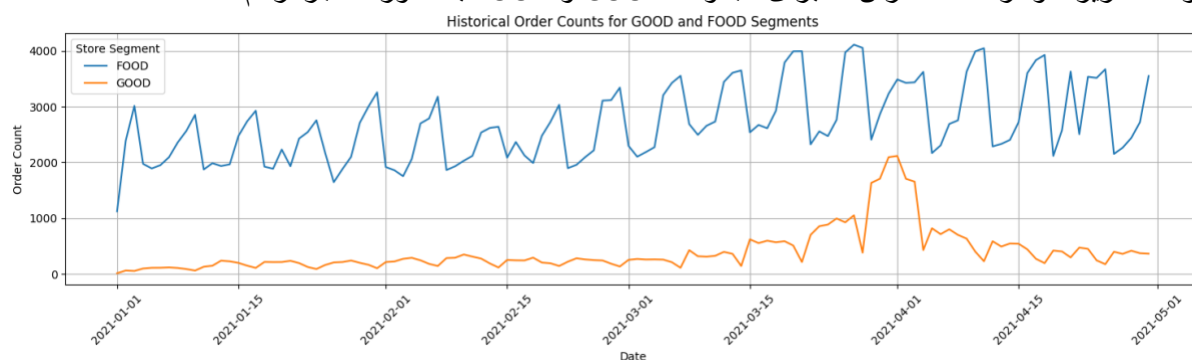


هیچ مقدار null در بخش store_segment وجود نداشت و برای بررسی تعداد GOOD و FOOD در سفارشات داریم:

```
store_segment
FOOD      321775
GOOD      47224
Name: count, dtype: int64
```

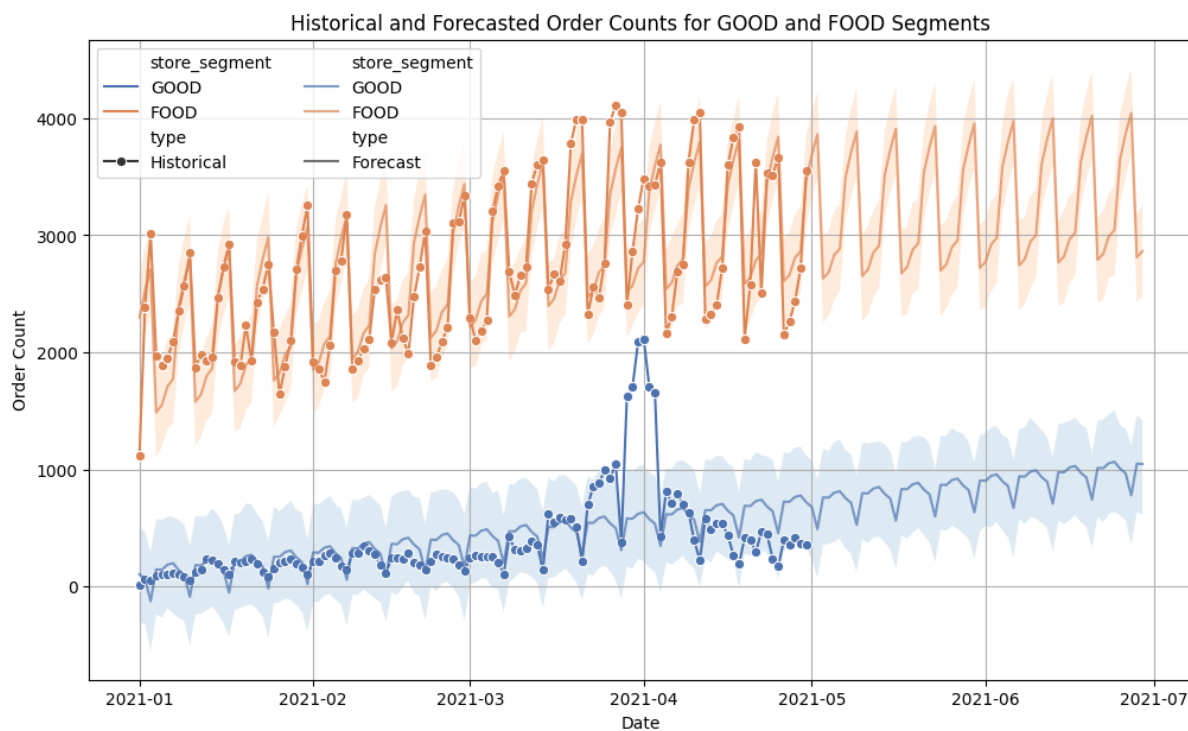
در سفارشات میزان FOOD حدودا 7 برابر GOOD بوده است.

در شکل زیر نمودار تعداد سفارش ها برای مجموعه GOOD و FOOD به صورت مجزا رسم شده است.



تاریخ 1-4-2021 (۱۲ فروردین ۱۴۰۰ آخرین روز تعطیلات سال نو) شامل قله در سفارشات بخش GOOD بودیم. اما بعد از آن روند آن به حالت قبلی باز گشته است. ۴ سیکل مربوط به ۴ هفته در هر ماه نوسان میزان را ملاحظه میکنید. طبق این شکل نیز تعداد سفارشات در بخش FOOD بسیار بیشتر حدودا ۷ برابر سفارشات GOOD بوده است.

برای پیش بینی روند دو ماه آینده از کتاب خانه Prophet استفاده شد.



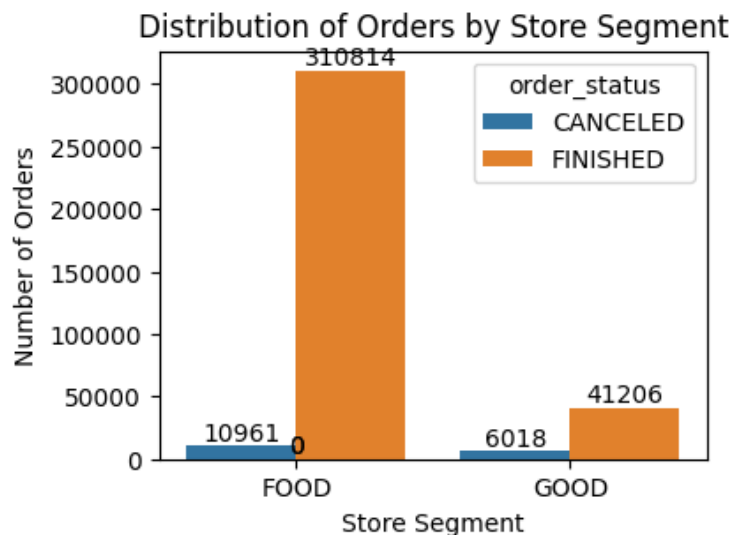
در این پیش بینی یک روند صعودی در تعداد سفارشات GOOD دیده میشود که ناشی از وجود قله در تاریخ 4-2021 است. بهتر است در تابع Prophet اطلاعات زمان آخر هفته ها و روزهای تعطیل داده شود تا دقیق تر پیش بینی انجام شود.

Task 3

برای بررسی توزیع سفارشات بعد merge دو جدول order و store از نمودار میله ای استفاده شد زیرا store- segment یک دیتای categorical است و این نمایش برای آن مناسب است.



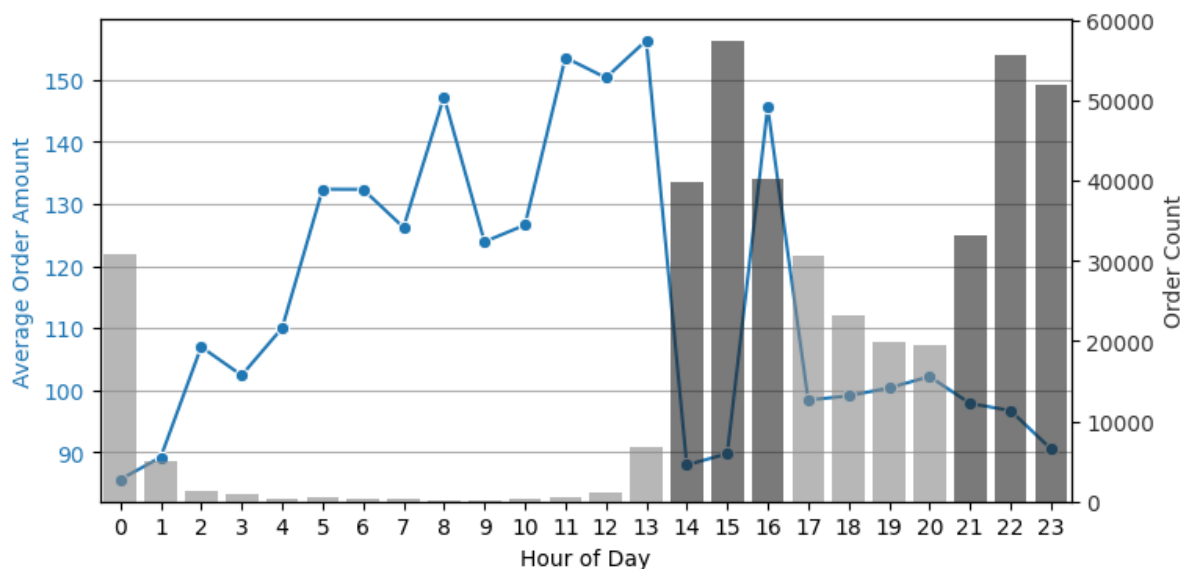
حدود 87 درصد سفارش ها FOOD بوده است. البته بخش از سفارش ها cancel شدند لذا شکل زیر را داریم:



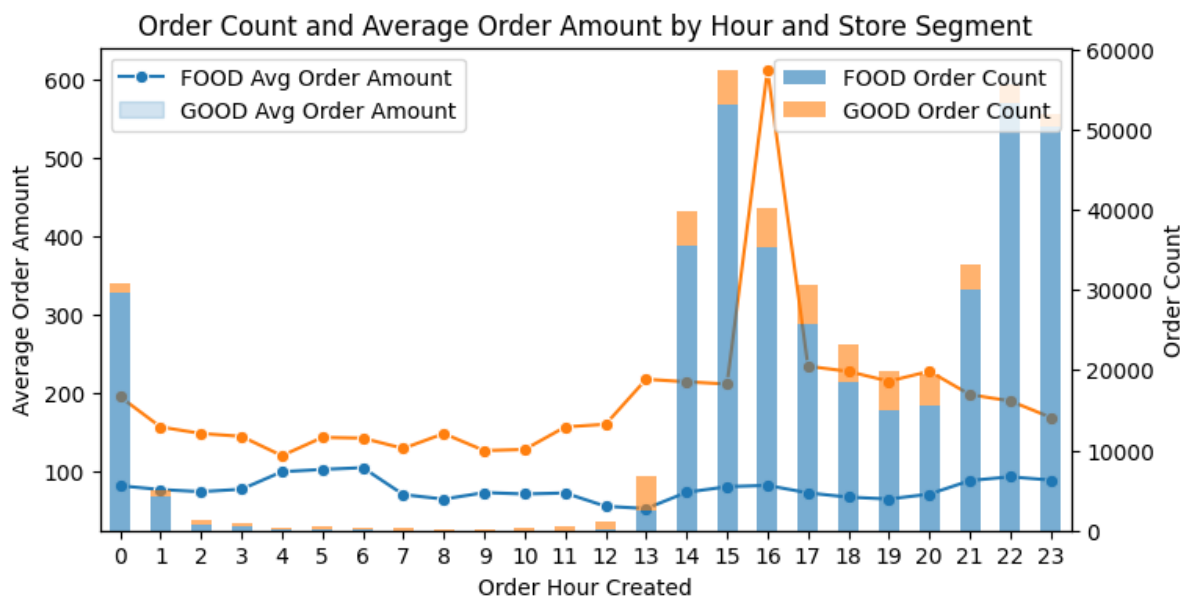
اگر سفارش هایی که cancel شدند را کنار بگذاریم باز هم مجموعه FOOD حدود 6 برابر GOOD بوده است. لازم به ذکر است در مجموعه GOOD حدودا ۱۳ درصد سفارش ها کنسل شده اند و در بخش FOOD تنها ۳ درصد سفارش ها کنسل شدند. لذا این بخش نیاز به بهبود دارد.

Task 4

برای محاسبه زمان اوج سفارش ها ابتدا واحد زمانی یک ساعته در نظر گرفته و با توجه به آن نمودار زیر را ملاحظه میکنید.

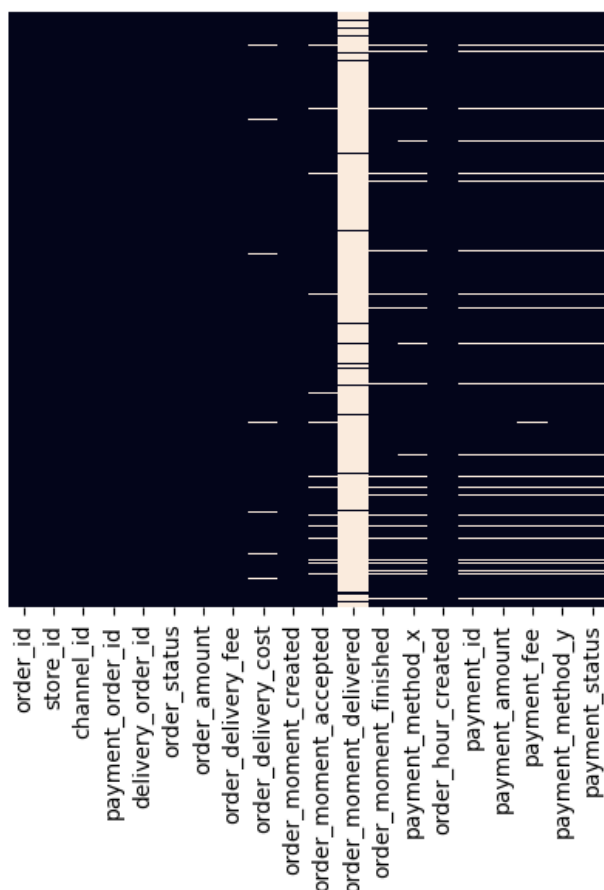


در این نمودار ساعت های 14,15,16,21,22,23 بیشترین تعداد سفارش ثبت شده است ($\text{quantile} > 0.75$) اما از طرفی میانگین order amount در این بازه ها کم بوده است. یک علت آن احتمالا میزان سفارش FOOD در مقابل GOOD که ارزان تر است احتمالا باشد لذا نمودار زیر را به تفکیک FOOD و GOOD رسم کردیم.



نتایج طبق انتظار نشان میدهد در ساعات شلوغی (ساعاتی که سفارش ها زیاد شده است) چون نسبت FOOD به GOOD زیاد شده متوسط میزان order amount پایین آمده است زیرا FOOD از GOOD بسیار قیمت پایین تری داشته. در ساعاتی چون 5 تا 12 GOOD بیشتر سفارش شده و لذا متوسط میزان order amount در این ساعات از ساعات شلوغی بسیار بالاتر بوده است.

برای بررسی رابطه متوسط سبد خرید (average order amount) در بازه های شلوغ با نوع پرداخت کاربر ابتدا جدول سفارش ها را با payments ادغام کردیم.



از لحاظ null بودن سفارش ها بخش `payment_method_x` و `payment_method_y` مهم است و این دو ستون با هم null شده اند ولی تعداد آنها زیاد نیست. و با توجه به شکل زیر تعداد category های `x` و `y` یکی است و

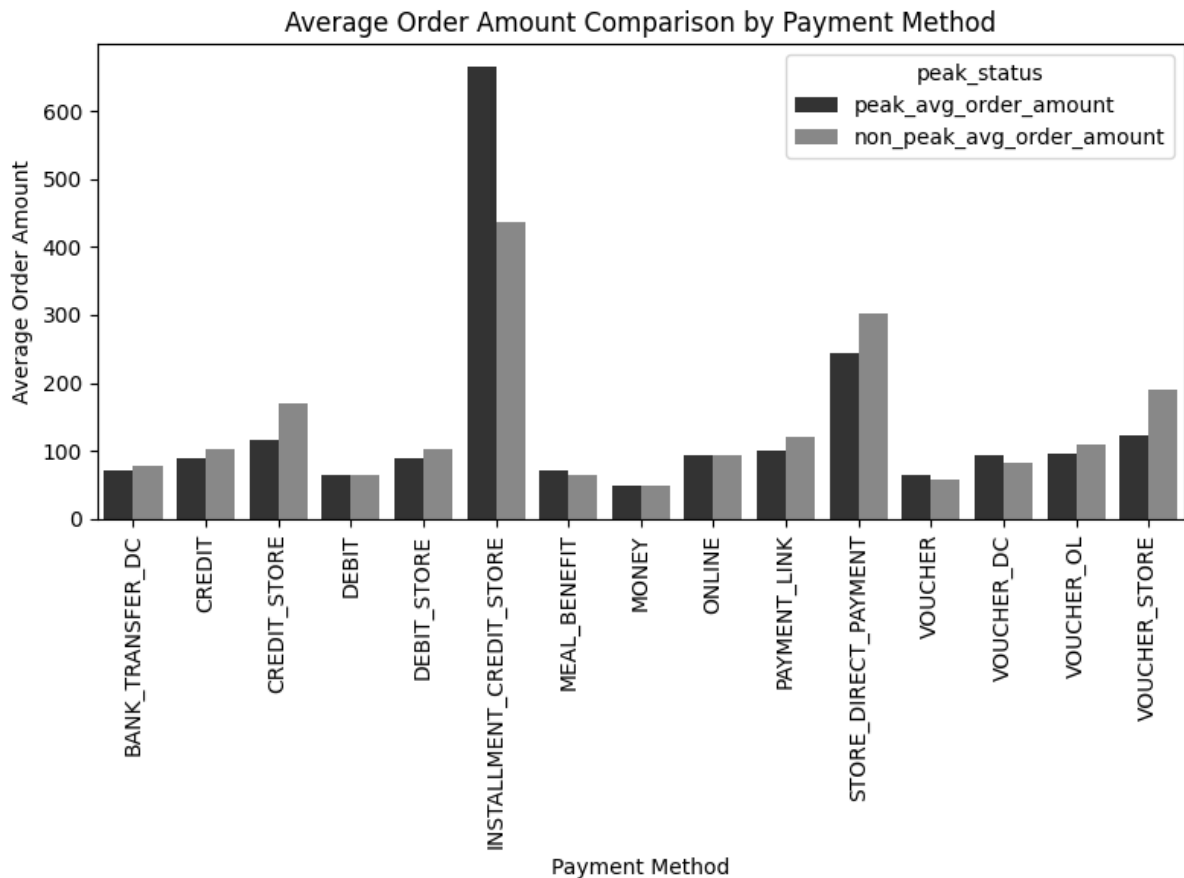
<code>merged_df["payment_method_x"].value_counts()</code>	<code>merged_df["payment_method_y"].value_counts()</code>
<code>payment_method_x</code>	<code>payment_method_y</code>
ONLINE 357982	ONLINE 357982
VOUCHER 99473	VOUCHER 99473
DEBIT 15331	DEBIT 15331
MEAL_BENEFIT 8213	MEAL_BENEFIT 8213
STORE_DIRECT_PAYMENT 6899	STORE_DIRECT_PAYMENT 6899
CREDIT 5942	CREDIT 5942
DEBIT_STORE 2548	DEBIT_STORE 2548
MONEY 2171	MONEY 2171
PAYMENT_LINK 1887	PAYMENT_LINK 1887
VOUCHER_STORE 1737	VOUCHER_STORE 1737
CREDIT_STORE 1542	CREDIT_STORE 1542
VOUCHER_DC 503	VOUCHER_DC 503
INSTALLMENT_CREDIT_STORE 473	INSTALLMENT_CREDIT_STORE 473
VOUCHER_OL 439	VOUCHER_OL 439
BANK_TRANSFER_DC 364	BANK_TRANSFER_DC 364
Name: count, dtype: int64	Name: count, dtype: int64

در جدول اصلی `payment` بخش `payment-order-id` گاهی تکرار دارد که برای یک `payment_order_id` یکسان بخش `payment_method_x` و `payment_method_y` آن متفاوت است. (این قسمت را متوجه نشدم!)

دو جدول جدید (برای زمان های شلوغ و زمان های خلوت) درست کرده که متوسط سبد خرید را بر اساس نوع پرداخت کاربر نشان میدهد.

	payment_method_x	peak_avg_order_amount	non_peak_avg_or	↑	↓	↔
0	BANK_TRANSFER_DC	71.090711	78.238960			
1	CREDIT	89.419572	101.480822			
2	CREDIT_STORE	116.627959	168.713523			
3	DEBIT	65.424041	65.315203			
4	DEBIT_STORE	90.094936	102.549549			
5	INSTALLMENT_CREDIT_STORE	665.093277	437.629730			
6	MEAL_BENEFIT	70.336511	63.905344			
7	MONEY	48.041728	47.677929			
8	ONLINE	93.225391	93.632424			
9	PAYMENT_LINK	101.212275	120.077885			
10	STORE_DIRECT_PAYMENT	244.527793	301.209326			
11	VOUCHER	64.142809	57.487616			
12	VOUCHER_DC	93.051517	81.395514			
13	VOUCHER_OL	94.874366	108.536593			
14	VOUCHER_STORE	123.350795	190.437631			

این تقسیم بندی را بر حسب payment_method_x انجام دادیم. در نهایت نمودار زیر مربوط به متوسط سبد خرید بر حسب نوع پرداخت به ازاء ساعت های شلوغ و ساعت های خلوت را نشان می دهد.

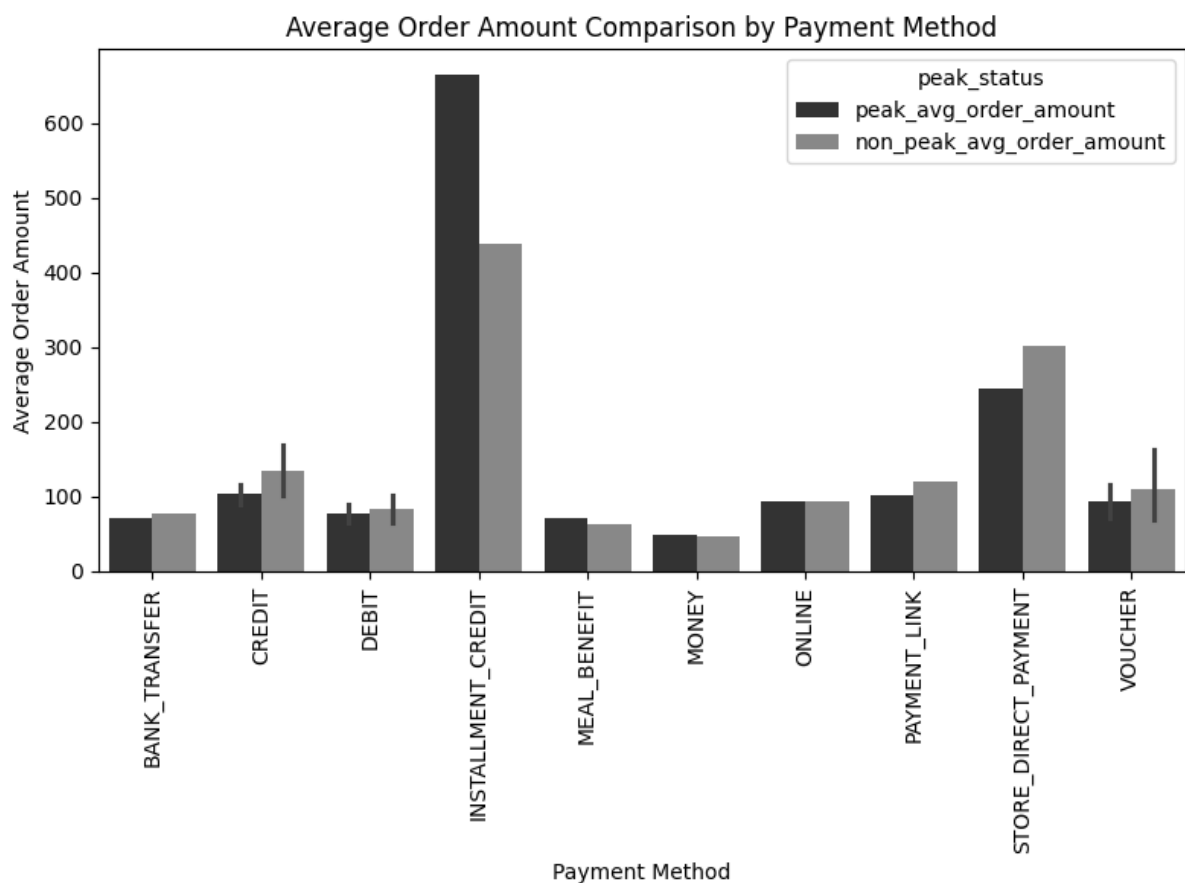


در ساعات شلوغ (بخش تیره رنگ نمودار) متوسط میزان سبد خرید در روش پرداخت **INSTALLMENT_CREDIT_STORE** بیشتر و بعد از آن روش **STORE_DIRECT_PAYMENT** بیشتر از بقیه روش ها بوده است.

با این حال با توجه به این که بعضی از این category ها بخش مشترکی دارند (به عنوان مثال **VOUCHER** در **VOUCHER_STORE** و **VOUCHER_OL** و **VOUCHER_DC** موجود است) می توان category دیگری ارائه داد و آن را نیز بررسی کرد.

برای این کار از map زیر استفاده کردیم.

```
mapping = {
  'ONLINE': 'ONLINE',
  'VOUCHER': 'VOUCHER',
  'DEBIT': 'DEBIT',
  'MEAL_BENEFIT': 'MEAL_BENEFIT',
  'STORE_DIRECT_PAYMENT': 'STORE_DIRECT_PAYMENT',
  'CREDIT': 'CREDIT',
  'DEBIT_STORE': 'DEBIT',
  'MONEY': 'MONEY',
  'PAYMENT_LINK': 'PAYMENT_LINK',
  'VOUCHER_STORE': 'VOUCHER',
  'CREDIT_STORE': 'CREDIT',
  'VOUCHER_DC': 'VOUCHER',
  'INSTALLMENT_CREDIT_STORE': 'INSTALLMENT_CREDIT',
  'VOUCHER_OL': 'VOUCHER',
  'BANK_TRANSFER_DC': 'BANK_TRANSFER'
}
```



طبق شکل بالا روش پرداخت `INSTALLMENT_CREDIT` و `STORE_DIRECT_PAYMENT` چه در ساعت های شلوغ و چه در ساعات خلوت متوسط سبد خرید بالا تری را داشته است. و پرداخت به روش `MONEY` کمترین میزان متوسط سبد خرید را دارد.