

به نام خدای رنگین کمان



دانشکده مهندسی صنایع

دانشگاه صنعتی شریف

## درس برنامه ریزی حمل و نقل

گزارش پروژه

استاد:

دکتر عرفان حسن نایبی

سروش اطمینان بخش (۹۸۱۰۳۸۹۱)

ماهان پورحسینی (۹۸۱۰۳۹۷۵)

ایمان شرکت بزازان (۹۸۱۰۴۱۹۹)

مهدی رحمانی طلب (۹۸۱۰۴۱۴۴)

نیم سال اول ۱۴۰۱-۰۲

## فهرست مطالب

۱.....	خواسته اول
۳.....	خواسته دوم
۴.....	خواسته سوم
۵.....	خواسته چهارم
۷.....	خواسته پنجم
۹.....	خواسته ششم
۱۰.....	خواسته هفتم
۱۲.....	خواسته هشتم
۱۴.....	خواسته نهم

## فهرست تصاویر

۱.....	شکل ۱: اطلاعات دیتاست
--------	-----------------------

## فهرست نمودارها

۷.....	نمودار ۱: تقاضا بر حسب ساعت
۱۴.....	نمودار ۲: میانگین قیمت سفر لغو شده توسط راننده بر حسب ساعت

## فهرست جداول

- جدول ۱: نوع (type) داده‌ها ..... ۲
- جدول ۲: درصد داده‌های Null ..... ۲
- جدول ۳: ماتریس OD تقاضای سفر ..... ۳
- جدول ۴: ماتریس OD سفرهای انجام‌شده ..... ۴
- جدول ۵: مجموع تقاضا (کل، انجام‌شده و برآورده نشده) ..... ۵
- جدول ۶: درصد تقاضاهای برآورده نشده ..... ۵
- جدول ۷: اطلاعات بیش‌ترین تقاضای برآورده‌نشده ..... ۶

## خواسته اول

پس از فراخوانی کتابخانه‌های موردنیاز و لود کردن داده، اطلاعات دیتاست را با دستور `data.info()` مشاهده می‌کنیم. داده‌ی ما ۲۶۷۶۸ سطر و ۱۱ ستون دارد. در جدول زیر اطلاعاتی مانند تعداد داده‌ی غیر Null در هر ستون و نوع داده‌های هر ستون قابل مشاهده است.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26768 entries, 0 to 26767
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   price_check_time      26768 non-null  object
1   passenger             26768 non-null  int64
2   origin                26768 non-null  int64
3   destination           26768 non-null  int64
4   price                 26768 non-null  int64
5   subsidy               26768 non-null  int64
6   distance              26507 non-null  float64
7   expected_duration     26498 non-null  float64
8   req_time              16717 non-null  object
9   driver                14403 non-null  float64
10  status                16717 non-null  float64
dtypes: float64(4), int64(5), object(2)
memory usage: 2.2+ MB
```

شکل ۱: اطلاعات دیتاست

نوع داده‌های هر ستون در جدول ۱ آورده شده است. چون ۴ ستون از جنس `float`، ۵ ستون `int` و ۲ ستون از جنس `datetime` داریم، نیازی به رمزگذاری یا انکودینگ طبقه‌بندی شده<sup>۱</sup> نداریم. به عبارتی داده‌های کیفی (مانند روزهای هفته) که قابل مقایسه عددی نیستند و نیاز به کمی کردن دارند در دیتاست ما وجود ندارد.

<sup>۱</sup> Categorical Encoding

Data Type	
price_check_time	object
passenger	int64
origin	int64
destination	int64
price	int64
subsidy	int64
distance	float64
expected_duration	float64
req_time	object
driver	float64
status	float64

جدول ۱: نوع (type) داده‌ها

در جدول ۲ ستون‌هایی که دارای داده‌ی Null هستند و درصد آن در کل ستون قابل مشاهده است.

	Attribute Name	Percentage of Null Values
0	distance	0.975045
1	expected_duration	1.008667
2	req_time	37.548565
3	driver	46.193216
4	status	37.548565

جدول ۲: درصد داده‌های Null

مشاهده می‌شود که دو ستون مسافت بر حسب متر<sup>۲</sup> و زمان انتظاری (تخمین) سفر بر حسب دقیقه<sup>۳</sup> دارای درصد مقادیر Null بسیار کمی هستند؛ بنابراین می‌توانیم این مقادیر Null را در هر ستون با میانگین آن ستون پر کنیم. این کار با دستور IterativeImputer از کتابخانه sklearn انجام شده است.

<sup>۲</sup> distance

<sup>۳</sup> expected\_duration

## خواسته دوم

برای استخراج ماتریس OD تقاضای سفر ابتدا یک دیتافریم خالی تشکیل می‌دهیم. سپس با یک for تودرتو، مقادیر سفر از تمامی مبدأها به تمامی مقاصد را از دیتاست داده‌شده فراخوانی کرده و وارد این دیتافریم می‌کنیم.

ماتریس OD تقاضای سفر در جدول ۳ آمده است.

**OD Matrix of Trip Requests**

	0	1	2	3	4
0	1422	1030	916	0	0
1	4096	3550	1535	1	0
2	4073	3538	1197	6	0
3	0	0	0	3712	724
4	0	0	0	626	342

جدول ۳: ماتریس OD تقاضای سفر

همان طور که از ماتریس OD فوق برمی‌آید، مبدأ با اندیس یک مبدأ بیشترین تعداد تقاضای سفر و منطقه با اندیس ۰ نیز مقصد بیشترین تعداد تقاضای سفر در مقایسه با سایر مناطق بوده است. همچنین کمترین تقاضای تولید سفر مربوط به منطقه با اندیس چهار و کمترین میزان جذب سفرهای تقاضاشده نیز مربوط به همین منطقه می‌باشد.

## خواسته سوم

برای استخراج ماتریس OD سفرهای انجام شده ابتدا یک دیتافریم خالی تشکیل می‌دهیم. سپس با یک for تودرتو، مقادیر سفرهای انجام شده از تمامی مبدأها به تمامی مقاصد را وارد این دیتافریم می‌کنیم. تفاوت این خواسته با خواسته قبل این است که هنگام وارد کردن داده‌ها به دیتافریم، تنها داده‌هایی وارد می‌شوند که ستون وضعیت درخواست مسافر<sup>۴</sup> آن‌ها برابر ۱ که معادل انجام سفر است، می‌باشد.

ماتریس OD سفرهای انجام شده در جدول ۴ آمده است.

OD Matrix of Done Trips					
	0	1	2	3	4
0	681	461	460	0	0
1	2040	1759	759	0	0
2	1940	1746	554	2	0
3	0	0	0	1656	340
4	0	0	0	268	172

جدول ۴: ماتریس OD سفرهای انجام شده

در ارتباط با این ماتریس نیز باید گفت که در بین سفرهای انجام شده منطقه با اندیس یک بیشترین مبدأ سفرها بوده و همچنین منطقه با اندیس صفر به عنوان بیشترین مقصد سفرها بوده است. ضمن این که منطقه‌ای که کمترین تعداد دفعات مبدأ و مقصد سفرها بوده، منطقه با اندیس شماره چهار می‌باشد.

<sup>4</sup> status

## خواسته چهارم

در جدول ۵ تعداد کل تقاضای سفر، سفرهای انجام شده و سفرهای برآورده نشده یا انجام نشده آورده شده است.

Demand		
	Type	Sum
0	Total	26768
1	Done	12838
2	Unfulfilled	13930

جدول ۵: مجموع تقاضا (کل، انجام شده و برآورده نشده)

از تقسیم تقاضای unfulfilled بر تقاضای Total، درصد تقاضای برآورده نشده به دست می آید. تقریباً ۵۲/۰۴ درصد تقاضای سفر برآورده نشده است. همچنین در جدول ۶ درصد تقاضاهایی که به سفر منجر نشده اند یا به عبارتی برآورده نشده اند برای همه مبداها و مقصدها نشان داده شده است.

Percentage of Requests Not Leading to Trips Per Each Origin and Destination					
	0	1	2	3	4
0	52.109705	55.242718	49.781659	nan	nan
1	50.195312	50.450704	50.553746	100.000000	nan
2	52.369261	50.650085	53.717627	66.666667	nan
3	nan	nan	nan	55.387931	53.038674
4	nan	nan	nan	57.188498	49.707602

جدول ۶: درصد تقاضاهای برآورده نشده

بیشترین درصد تقاضای برآورده نشده مربوط به سفرهایی است که از مبدأ ۱ به مقصد ۳ صورت می گیرند و طبق جدول ۷، ۱۰۰ درصد تقاضای این جفت مبدأ - مقصد برآورده نمی شود. در واقع طبق نتایج بین مبدأ ۱ و مقصد ۳ سفری صورت نگرفته است.

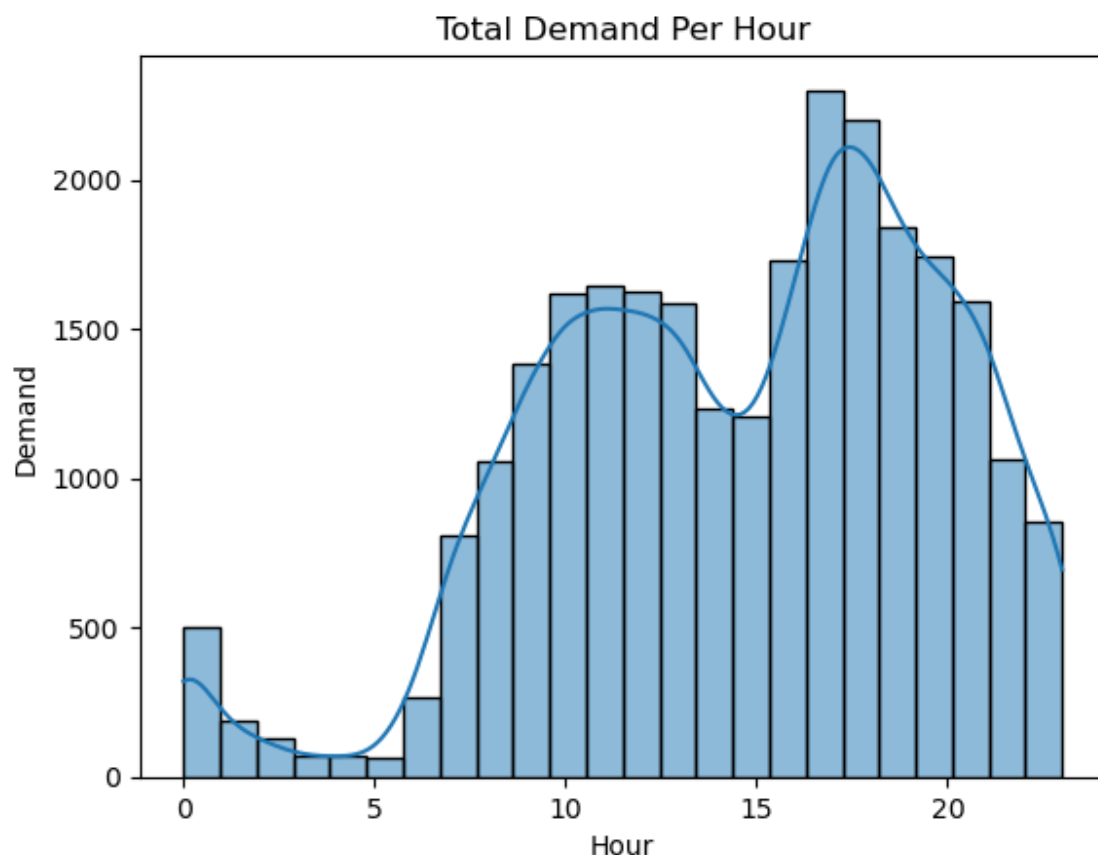


0	Origin (From)	1
1	Destination (To)	3
2	Unfulfilled Percentage	100.0

جدول ۷: اطلاعات بیشترین تقاضای برآورده نشده

## خواسته پنجم

نمودار ۱ نشان‌دهنده کل تقاضای سفر در بازه‌های ۱ ساعته در طول شبانه‌روز است.



نمودار ۱: تقاضا بر حسب ساعت

همان طور که از نمودار رسم شده مشخص است کمترین میزان تقاضای سفر در حد فاصل ساعت نیمه شب تا پنج الی شش صبح رخ می‌دهد که کاملاً منطقی است؛ پس از این ساعات با شروع ساعت کاری و ایجاد جریان کاری و زندگی روزمره در سطح شهر میزان تقاضای سفر به مرور افزایش می‌یابد.

این افزای تا ساعت ده الی یازده صبح ادامه پیدا می‌کند و از آنجا به بعد نرخ نزولی پیدا می‌کند و مجدداً از ساعت سه بعد از ظهر نرخ تقاضای سفر افزایش می‌یابد به طوری که در حد فاصل ساعت چهار تا پنج عصر به بیشترین مقدار خود در طول شبانه‌روز می‌رسد. پس از گذر از این ساعات پیک تقاضای سفر مجدد نرخ تقاضا به مرور کاهش می‌یابد.

به طور کلی نمودار رسم شده یک روند کاملاً منطقی از زندگی روزمره و اتفاقات آن را شامل می‌شود؛ در ساعات ابتدایی روز میزان تقاضا به دلیل شروع فعالیت‌ها افزایش می‌یابد و این نرخ افزایش تا ساعاتی قبل از ظهر ادامه پیدا می‌کند. در ساعات بعد از ظهر روز نیز چون به طور مثال عموم ادارات، بانک‌ها و ... تعطیل شده‌اند، حجم

تقاضا نیز مجدد افزایش می‌یابد. از طرفی در این ساعت از روز بخشی از جابه‌جایی‌ها به منظور تفریح، خرید و ... توسط خانوارها و افراد اتفاق می‌افتد که خود نیز عاملی برای افزایش تقاضای سفر است که تمامی این عوامل در کنار یکدیگر سبب می‌شود در حد فاصل ساعت چهار تا پنج بعد از ظهر به حداکثر تقاضای سفر برسیم.

در نتیجه باید دقت داشت با توجه به حجم بالای حمل‌ونقل و تقاضای سفر در ساعات منتهی به بعدازظهر تعیین مناسب قیمت‌ها بسیار حائز اهمیت می‌باشد که در صورت انجام درست این مورد می‌توان حجم زیادی از تقاضای آن بخش از ساعات روز را به خود اختصاص داد و سهم خود را از بازار بالا برد.

## خواسته ششم

در این قسمت به بررسی تأثیر قیمت پیشنهادی در لغو سفر از سمت مسافر می‌پردازیم که آیا این تأثیر معنادار بوده و یا خیر، تأثیر چندانی در لغو سفر از سمت مسافر نداشته است؟! در واقع به دنبال بررسی پراکندگی جامعه آماری این دو پارامتر نسبت به یکدیگر هستیم. برای این بررسی وضعیت درخواست مسافران به دو دسته تقسیم‌بندی می‌شوند: دسته اول statusهای دو که سفرهای لغو شده توسط مسافر است و سایر statusها به جز blank که در نظر گرفته نمی‌شود زیرا که هیچ درخواستی ثبت نشده که حال بخواهد تأثیر قیمت بر روی آن بررسی شود.

به همین منظور در اینجا به کمک تست فیشر به بررسی این فرض می‌پردازیم با در نظر گرفتن این مسئله که فرض صفر ما عدم وجود هم‌بستگی بین قیمت پیشنهادی و لغو سفر از طرف مشتری است و به تبع فرض یک وجود هم‌بستگی معنادار بین قیمت و لغو سفر از طرف مشتری می‌باشد.

در ارتباط با تست فیشر می‌دانیم که هر چه قدر مقدار  $F$  بیشتر باشد برای ما بهتر است؛ از طرفی اگر مقدار  $f$  محاسبه شده از  $F$  کمتر باشد ما می‌توانیم فرض صفر را رد کنیم. در اینجا پس از انجام محاسبات مقدار  $f$  برابر با ۰.۴۱۴ و مقدار  $F$  نیز برابر با ۳.۸۴۲ می‌شود. مشخص است که  $f$  از  $F$  کمتر است و همین مسئله برای رد کردن فرض صفر کافی است و می‌توان دریافت که یک هم‌بستگی و به اصطلاح correlation معنادار بین قیمت پیشنهادی و لغو سفر از طرف مسافر وجود داشته است.

همچنین اگر از کای تست برای بررسی این قسمت استفاده کنیم این طور برمی‌آید که مقدار  $P$  بدست آمده از ۰.۰۵ کمتر است در نتیجه مجدداً می‌توان فرض صفر را رد کرد و به این نتیجه رسید که رابطه و هم‌بستگی معناداری بین قیمت پیشنهادی و لغو سفر از طرف مسافر وجود دارد.

## خواسته هفتم

این بار و در این خواسته از مسئله به بررسی رفتار رانندگان در تعامل با قیمت پیشنهادی می‌پردازیم. در خواسته شماره شش به کمک تست فیشر، این نتیجه حاصل شد که قیمت پیشنهادی به مسافران به ازای مسافت‌ها و سفرهای گوناگون تأثیر معناداری در لغو سفر از سمت آن‌ها داشته است. حال به بررسی تأثیر این عامل در لغو سفر از سمت رانندگان در سطح اطمینان ۹۵٪ می‌پردازیم. در این قسمت نیز باید داده‌ها را به منظور مقایسه دسته‌بندی کرد؛ رفتار رانندگان در مقایسه با رفتار مسافران مقداری متفاوت است به همین منظور نیاز است که دو تحلیل متفاوت ارائه شود. در ابتدا قسمت blank را کنار گذاشته و یک دسته را status3 یعنی لغو سفر توسط راننده و دسته بعدی را سایر statusها در نظر می‌گیریم و در تحلیل بعدی status3 و status4 را با هم در نظر گرفته با سایر statusها مقایسه می‌کنیم.

به همین منظور برای این قسمت نیز از تست فیشر استفاده می‌کنیم و توضیحات خواسته قبلی، این بار در ارتباط با رانندگان مطرح است. برای محاسبه تست فیشر مجدداً به محاسبه آماره‌های  $f$  و  $F$  از جوامع آماری مرتبط می‌پردازیم که این بار مقدار  $f$  برابر با ۰.۶۸۹ و مقدار  $F$  یا همان  $F$  Critical Value برابر با ۳.۸۴۲ خواهد شد.

کاملاً مشخص است که مقدار بحرانی  $F$  در هر دو خواسته شش و هفت با هم برابر خواهد بود زیرا که در هر دو قسمت اثر قیمت پیشنهادی در حال بررسی است و همین موضوع سبب برابر و یکی شدن مقدار بحرانی  $F$  در هر دو تست می‌شود.

با توجه به مقادیر به دست آمده برای  $f$  و  $F$  و کوچک‌تر بودن مقدار  $f$  از  $F$  مجدداً این نتیجه حاصل می‌شود که قیمت پیشنهادی با لغو سفر از طرف رانندگان هم‌بستگی و correlation معناداری دارد.

در این قسمت نیز مجدداً از کای تست برای بررسی دوباره استفاده می‌کنیم که به کمک این تست نیز اینطور نتیجه می‌شود که لغو سفر از طرف رانندگان هم‌بستگی معناداری با قیمت پیشنهادی دارد.

به طور کلی پس از بررسی خواسته‌های شش و هفت می‌توان این را دریافت که قیمت پیشنهادی تأثیر بسزایی در لغو سفرها از سمت مسافران و یا رانندگان دارد.

این تأثیر به این شکل است که عموم مسافران با قیمت‌هایی روبرو می‌شوند که از نظر آن‌ها برای مسیر مد نظری که برای سفر خود انتخاب کرده‌اند، بسیار گران است و مسافت مسیر و مدت زمانی که سفر طول می‌کشد، ارزش آن مقدار هزینه را ندارد و همین سبب می‌شود که مسافر از تقاضای سفر خود صرف نظر کند و آن را لغو کند.

اما اگر از دید رانندگان به ماجرا نگاه کنیم به تحلیل مشابهی خواهیم رسید؛ از نظر رانندگان نیز سفرهای پیشنهادی بعضاً قیمت‌هایی دارند که با توجه مسافت، زمان سفر و دیگر عوامل به صرفه نیست و توقع قیمت و در نتیجه دریافتی بیشتری را از همان مسیر دارند لذا تمایل چندانی به انجام و پذیرش آن سفر با قیمت پیشنهادی برنامه ندارند و آن را کنسل می‌کنند.

اما نکته مهم در بررسی خواسته‌های این دو قسمت توجه به نرمال بودن داده‌هاست زیرا که اگر داده‌ها نرمال نباشند امکان استفاده از تست‌های قبلی نمی‌باشد و در اینجا نیز داده‌ها از توزیع نرمال پیروی نمی‌کنند به همین منظور از کای تست استفاده کردیم که دیگر مشکل نرمال‌سازی را نداشته باشیم.

## خواسته هشتم

یکی دیگر از خواسته‌هایی که در بخش تحلیل رفتار رانندگان به بررسی آن پرداخته شده است، احتمال رد درخواست مسافر توسط راننده در ساعت‌های پیک و غیر پیک و بررسی وجود یا عدم وجود تفاوت معنادار بین این احتمال می‌باشد.

در این قسمت نیز باید به بررسی رابطه بین رد سفرها توسط راننده و همچنین ساعت تقاضای سفر که آیا در ساعات پیک هستیم یا خیر پردازیم. به همین منظور از تست `norm.ppf` استفاده می‌کنیم که کدهای مربوط به آن در فایل نیز قابل مشاهده است. در این قسمت اگر سفرهایی که در ساعات پیک و غیرپیک انجام شده را به ترتیب با  $PA$  و  $NA$  مشخص کنیم و همچنین احتمالات زیر را بدست بیاوریم:

$$p_1 = 1 - \frac{\sum PA + \sum NA}{N_{PA} + N_{NA}}$$

$$p_2 = 1 - \frac{\sum PA}{\text{کل تقاضای سفرهای در ساعات پیک}}$$

$$p_3 = 1 - \frac{\sum NA}{\text{کل تقاضای سفرهای در ساعات غیرپیک}}$$

حال به کمک تست‌های موجود مقادیر زیر را محاسبه می‌کنیم:

$$Z = \frac{p_2 - p_3}{\sqrt{\frac{p_1(1 - p_1)}{\text{کل تقاضای سفرهای در ساعات پیک}}}}$$

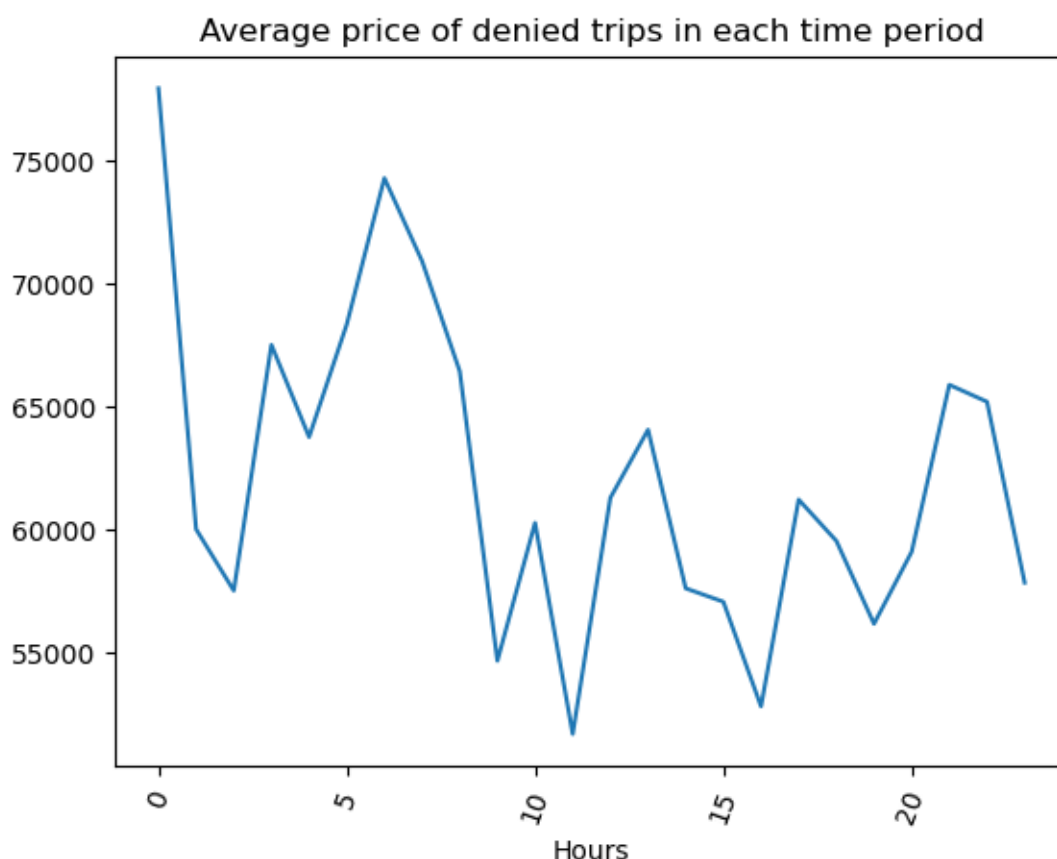
همچنین مقدار بحرانی آزمون در سطح اطمینان ۹۵٪ را محاسبه کرده و  $n$  می‌نامیم اگر قدرمطلق  $Z$  از قدر مطلق  $n$  کمتر باشد آن‌گاه فرض صفر رد پذیرفته می‌شود و هیچ رابطه‌ای بین رد کردن سفر و ساعت تقاضای سفر وجود ندارد اما با توجه به این که مقدار  $Z$  برابر ۳.۵۰۱ و مقدار  $n$  برابر با ۰.۰۰۳۶ شده است در نتیجه فرض صفر رد شده و می‌توان دریافت که احتمال رد و کنسل کردن سفر تحت تأثیر بودن یا نبودن در ساعات پیک می‌باشد و اختلاف معناداری بین احتمال رد سفر در ساعات غیر پیک و احتمال رد سفر در ساعات پیک می‌باشد.

نتیجه بدست آمده نیز با شهود کلی در این زمینه همخوانی دارد؛ در ساعات پیک به دلایل مختلف از جمله شلوغی مسیرها و ترافیک سنگین و در نتیجه آن مدت زمان زیاد سفر احتمال لغو و رد سفرها توسط رانندگان بسیار بیشتر است و این مسئله نیز در این قسمت خودش را نشان داده است.



## خواسته نهم

در آخرین بخش از تحلیل رفتار رانندگان به بررسی تأثیر میانگین قیمت در ساعات مختلف روز بر روی رد درخواست‌های سفر توسط رانندگان پرداخته می‌شود. این قسمت روند تحلیل را به صورت نموداری مد نظر دارد و به همین منظور نمودار میانگین قیمت سفرهای رد شده توسط رانندگان را بر حسب ساعات مختلف روز رسم می‌کنیم که نمودار خروجی به شکل زیر می‌باشد.



نمودار ۲: میانگین قیمت سفر لغو شده توسط راننده بر حسب ساعت

همان طور که مشاهده می‌شود در ساعات نیمه شب قیمت‌ها بسیار بالاست ولی بنابه دلایل مختلفی راننده تصمیم به لغو و رد سفر می‌گیرد؛ دلایلی مانند مسافت‌های بسیار دور و بحث امنیت و سایر دلایل. در ادامه قیمت‌ها روند کاهشی پیدا می‌کنند. در ادامه و در ساعات آغازین روز با توجه به این که حجم حمل و نقل در سطح شهر و تقاضاهای سفر بالا می‌رود و از طرفی عموم افراد دارای عجله هستند، قیمت‌ها نیز بالا می‌رود اما رانندگان ممکن است به دلایل مختلف مانند ترافیک بیش از اندازه و مسافت‌های طولانی سفرها را رد کنند.

در ادامه روز نیز به طور کلی سطح قیمت‌ها کاهش یافته و با یک روند سینوسی مواجه است که در موارد قیمت افزایش پیدا کرده و مجدداً کاهش یافته و دوباره افزایش و سپس کاهش و ... اما به وضوح مشخص است

که از ساعت هشت صبح به بعد تقریباً یک شیفت و جابه‌جایی اساسی در سطح قیمت‌ها رخ داده است و میانگین قیمت سفرهایی که رد شده‌اند به شدت کاهش پیدا کرده است که خود بیانگر این است که با فاصله گرفتن از ساعات ابتدایی روز، قیمت‌های پایین، عامل اصلی رد سفرها توسط رانندگان می‌باشد.