

سوالات تحلیلی

سوال ۱:

با توجه به این که هدف اصلی این جشنواره جذب مشتری جدید و افزایش آگاهی از برند ماهکس است، می بایست یک یا چند شاخص KPI (بلند مدت و کوتاه مدت) مناسب به منظور سنجش میزان اثر بخشی این جشنواره انتخاب شود. از طرفی این ۲۰٪ تخفیف نوعی سرمایه گذاری محسوب می شود، بنابراین می بایست میزان بازگشت سرمایه در این شاخص در نظر گرفته شود. این شاخص ها باید بتوانند به سوالات زیر پاسخ دهند.

- میزان افزایش مشتریان جدید (به عبارتی نرخ ورودی) در طی این ۳ ماه نسبت به سال قبل چقدر است؟ (کوتاه مدت)
 - چه تعداد از این مشتریان بعد از پایان این جشنواره سفارش جدیدی می دهند؟ (نرخ تبدیل) (بلند مدت)
 - به ازاء هر مشتری جدید چه مقدار هزینه شده؟
 - به ازاء هر مشتری جدید چه مقدار درآمد کسب شده؟ (به عبارتی هر مشتری جدید چقدر برای ما ارزش دارد؟)
- می بایست داده ها و شاخص های مورد نیاز برای محاسبه موارد فوق ثبت و نظارت شود (داده های مربوط به مشتریان، سفارشات و پرداخت های آنان)، بدین منظور نیاز به یک سیستم ردیابی مشتری احساس می شود.

سوال ۲:

به منظور کاهش هزینه های نگهداری و تعمیرات یکی از اهدافی که می توان دنبال کرد کاهش اصطهلاک وسایل نقلیه در دسترس است، بدین منظور اقدامات زیر میتواند صورت گیرد:

- تاسیس هاب های توزیع: به دلیل وسعت تهران، تاسیس هاب های توزیع میتواند به طور چشم گیری مسافت پیموده شده توسط خودرو ها را کاهش دهد، مکانیابی و تخصیص این مراکز باید به صورت بهینه و اصولی انجام گیرد.
- تخصیص بهتر نقاط دریافت و تحویل بسته ها با استفاده از تکنیک های یادگیری ماشین بدون نظارت
- اجرای نت پیشگیرانه و سرویس دوره ای وسایل نقلیه می تواند از خرابی های اساسی و اصطهلاک قطعات دیگر خودرو جلوگیری کند.
- اجرای نت پیشبینانه و پیشبینی زمان خرابی قطعات به خصوص (با نظر کارشناس) خودرو و در نتیجه تعیین زمان بهینه تعویض آنها می توان هزینه های نت را کاهش داد.

سوالات فنی

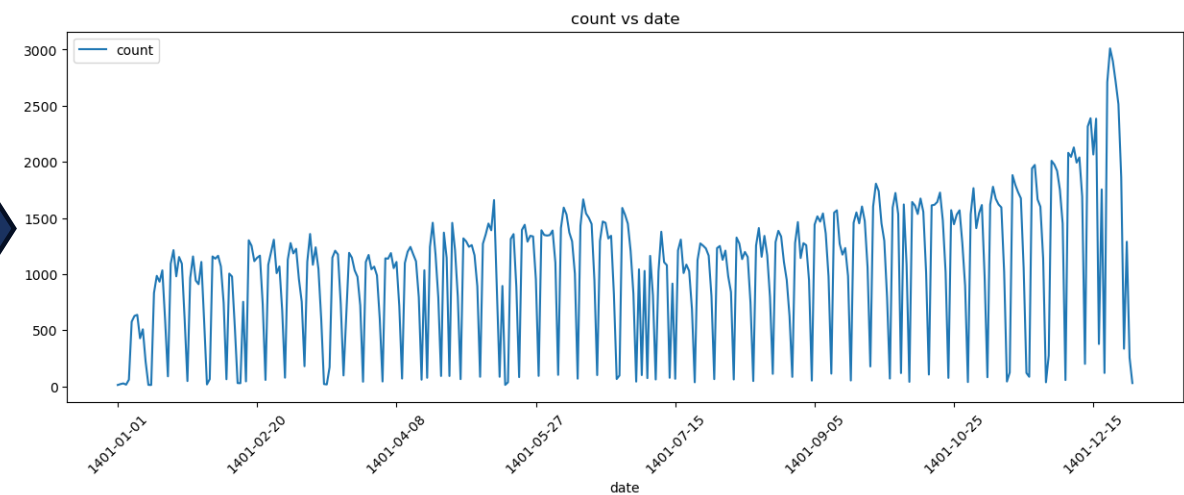
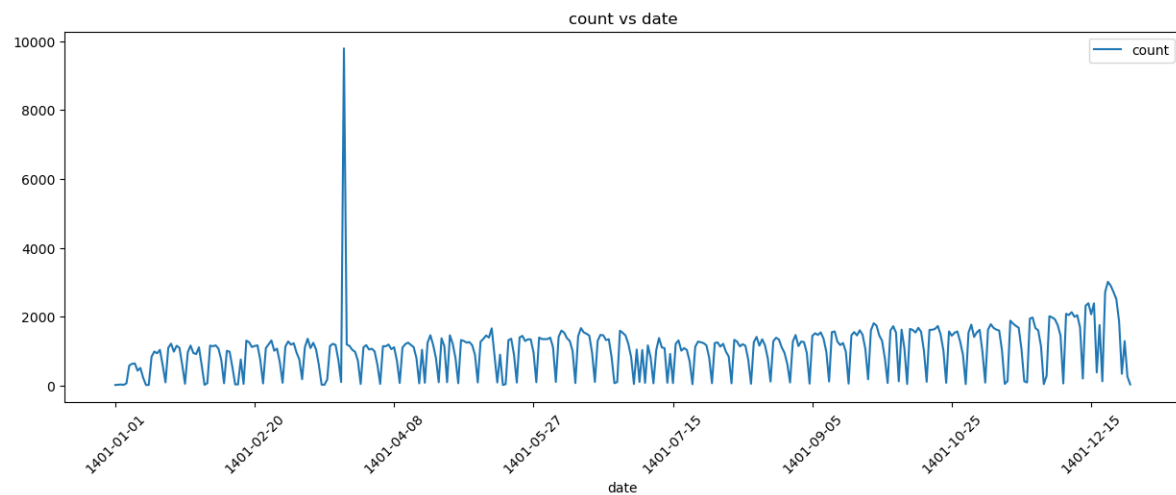
سوال ۱:

شماره سفارش	تاریخ ایجاد سفارش	وزن (کیلوگرم)	استان فرستنده	شهر فرستنده	استان گیرنده	شهر گیرنده
384259	1401-10-01	1.0	کردستان	بانه	اصفهان	کاشان
384260	1401-10-01	18.0	کردستان	مریوان	تهران	تهران
384261	1401-10-01	3.0	کردستان	مریوان	تهران	تهران
384262	1401-10-01	3.0	آذربایجان غربی	ارومیه	تهران	تهران
384263	1401-10-01	0.5	خوزستان	اهواز	تهران	تهران

داده های ارائه شده دارای ۳۸۴۲۶۴ ردیف و به فرمت فوق است، از آنجایی که متغیر پاسخ ما تعداد مرسولات است می بایست فرمت آن را به صورت زیر تغییر دهیم. تعداد بسته های ارسالی در هر روز قابل مشاهده است.

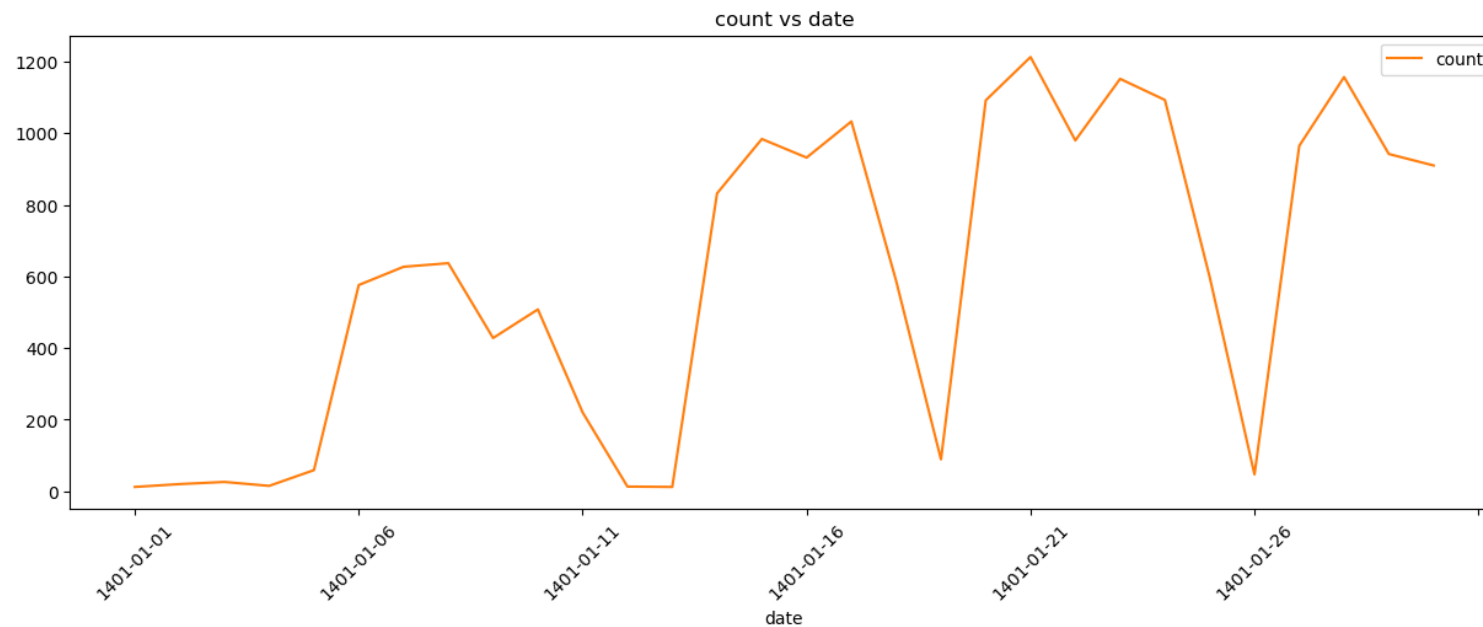
count	
date	
1401-01-01	12
1401-01-02	20
1401-01-03	26
1401-01-04	15
1401-01-05	59

داده پرت:



■ مصور سازی داده های خام: به نظر می رسد یک مورد داده پرت داریم، می بایست دلیل وجود این داده پرت بررسی شود. اما فعلا آن را با میانگین دو داده قبل و بعد آن جایگزین می کنیم.

فصلی بودن داده ها:



■ داده ها به صورت هفتگی، فصلی هستند، بررسی این که آیا این داده ها به صورت سالیانه، فصلی هستند ممکن نیست چرا که تنها حجم یک سال داده را در اختیار داریم.

پردازش داده ها:

	count	day_of_month	month	day_of_week	week
date					
1401-01-01	12	1	1	2	1
1401-01-02	20	2	1	3	1
1401-01-03	26	3	1	4	1
1401-01-04	15	4	1	5	1
1401-01-05	59	5	1	6	1
1401-01-06	576	6	1	0	2
1401-01-07	627	7	1	1	2
1401-01-08	637	8	1	2	2
1401-01-09	428	9	1	3	2
1401-01-10	508	10	1	4	2

	count	Target-27	Target-26	Target-25	Target-24	Target-23	Target-22	Target-21	Target-20	Target-19	...	Target-10	Target-9	Target-8	Target-7
date															
1401-01-28	1157	12	20	26	15	59	576	627	637	428	...	587	89	1092	1213
1401-01-29	942	20	26	15	59	576	627	637	428	508	...	89	1092	1213	980
1401-01-30	910	26	15	59	576	627	637	428	508	220	...	1092	1213	980	1152
1401-01-31	1108	15	59	576	627	637	428	508	220	13	...	1213	980	1152	1093
1401-02-01	577	59	576	627	637	428	508	220	13	12	...	980	1152	1093	598

	count	day_of_month	month	day_of_week	week	Target-27	Target-26	Target-25	Target-24	Target-23	...	Target-10
date												
1401-01-28	1157	28	1	1	5	12.0	20.0	26.0	15.0	59.0	...	587.0
1401-01-29	942	29	1	2	5	20.0	26.0	15.0	59.0	576.0	...	89.0
1401-01-30	910	30	1	3	5	26.0	15.0	59.0	576.0	627.0	...	1092.0
1401-01-31	1108	31	1	4	5	15.0	59.0	576.0	627.0	637.0	...	1213.0
1401-02-01	577	1	2	2	5	59.0	576.0	627.0	637.0	428.0	...	980.0

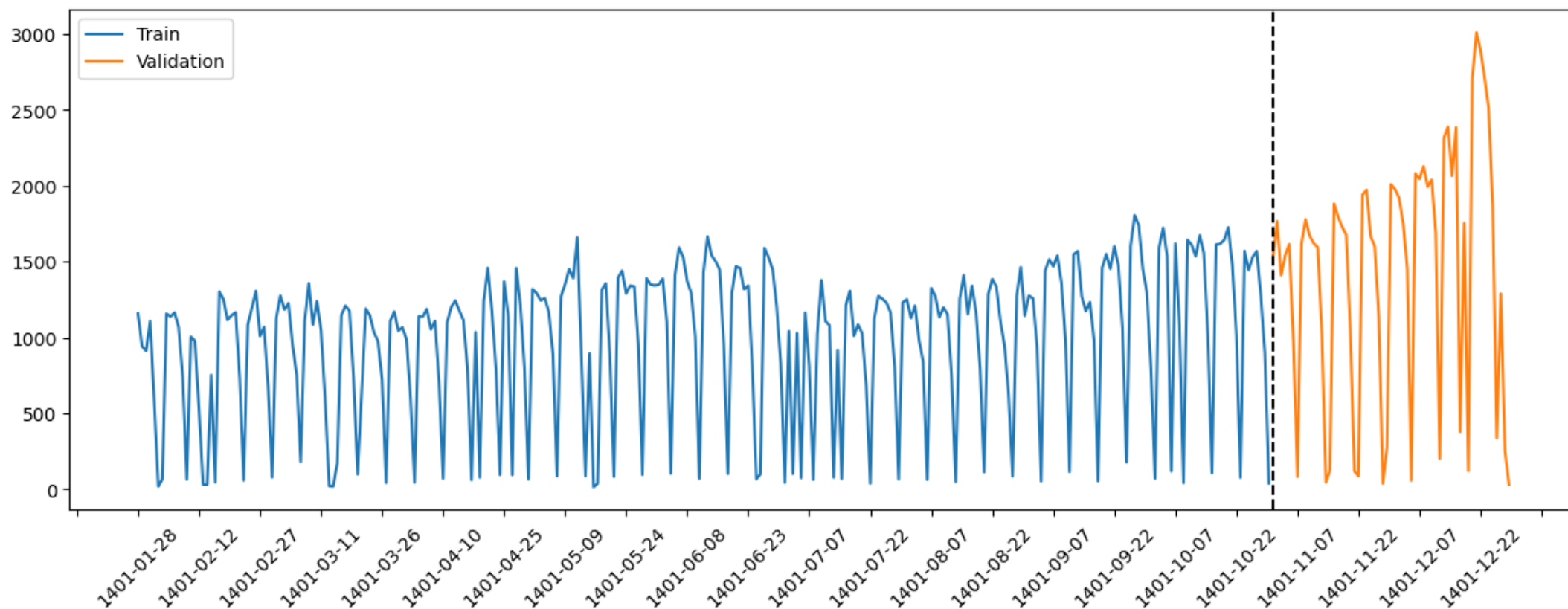
به سه شکل داده ها را پردازش می کنیم:

۱- ساخت فیچر های شماره روز از ماه، روز از هفته، ماه و هفته

۲- ساخت دیتاست سلسله مراتبی

۳- ترکیب هر دو

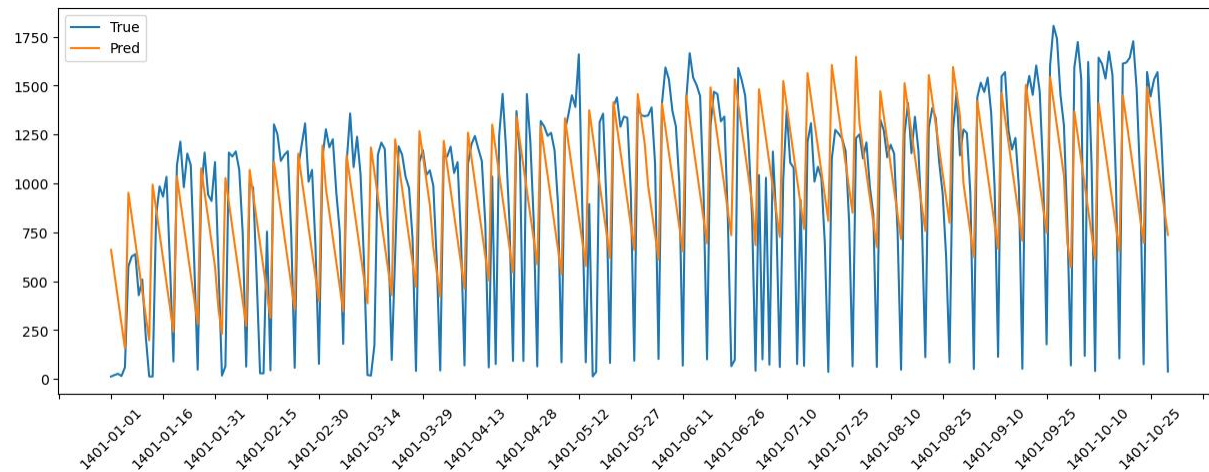
■ داده های دو ماه آخر را به عنوان داده های Cross Validation انتخاب می کنیم.



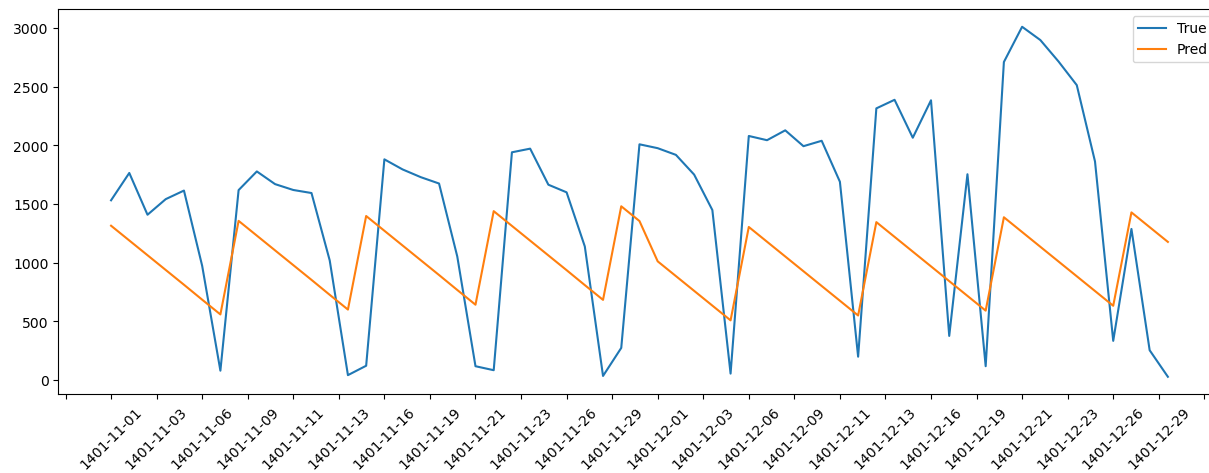
مدل سازی

- ابتدا با استفاده از سه دیتاست مذکور اقدام به آموزش سه مدل رگرسیون، گرادیان تقویتی و شبکه عصبی LSTM کرده و با استفاده از دیتاست validation اقدام به اعتبار سنجی نتایج مدل ها پرداخته و سپس با استفاده از بهترین مدل ۳ ماه اول سال ۱۴۰۲ را پیشبینی می کنیم.

رگرسیون

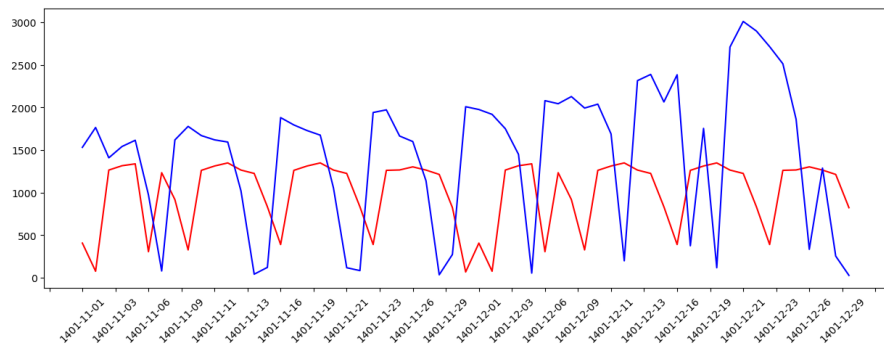


مقادیر واقعی و پیشبینی شده داده های Train



مقادیر واقعی و پیشبینی شده داده های validation با
 $rmse = 90.1$

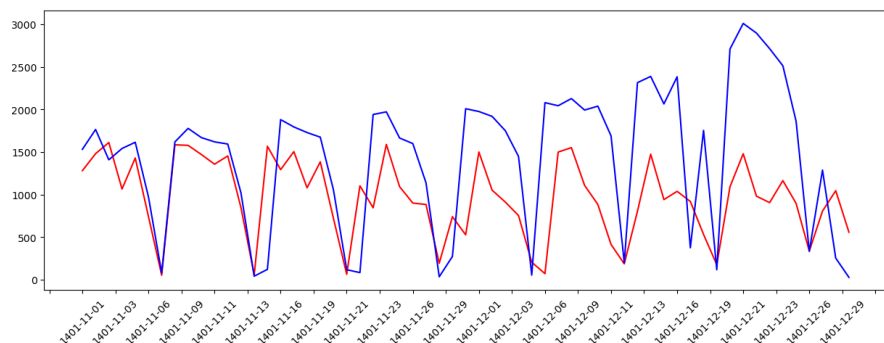
XGBOOST



دیتاست نوع ۱

مقادیر واقعی و پیشبینی شده داده های validation با

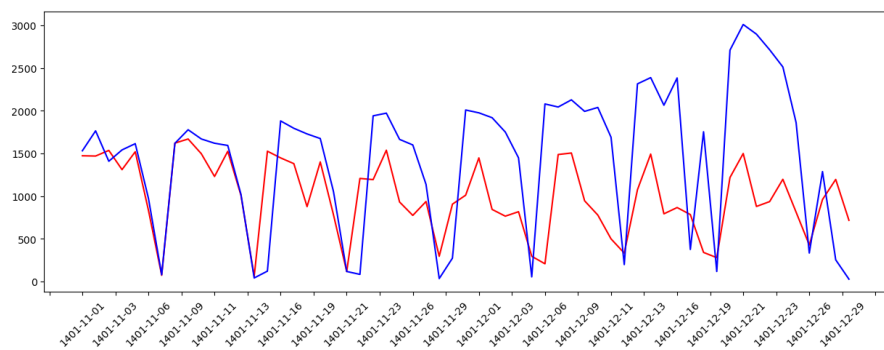
$$11.3/4 = \text{rmse}$$



دیتاست نوع ۲

مقادیر واقعی و پیشبینی شده داده های validation با

$$862/1 = \text{rmse}$$

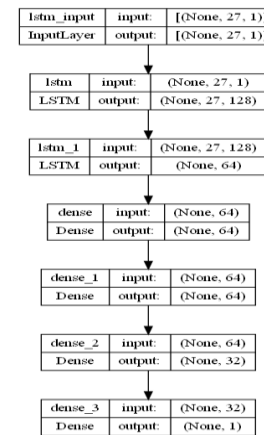
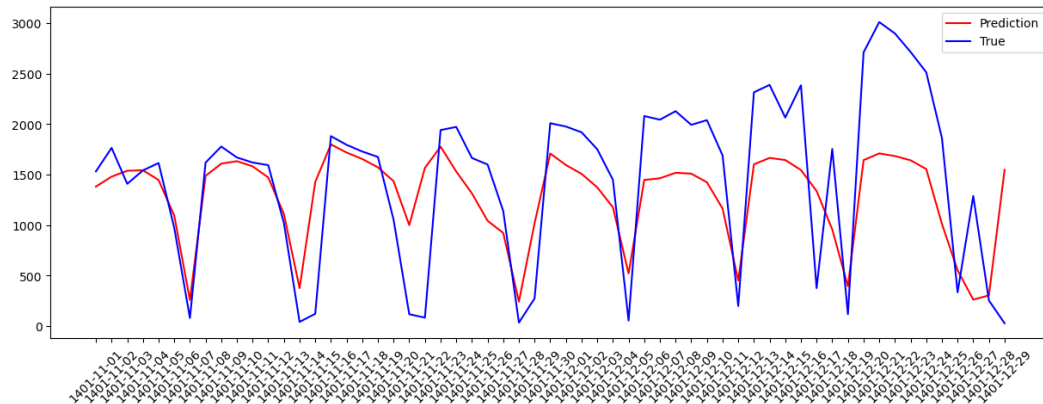
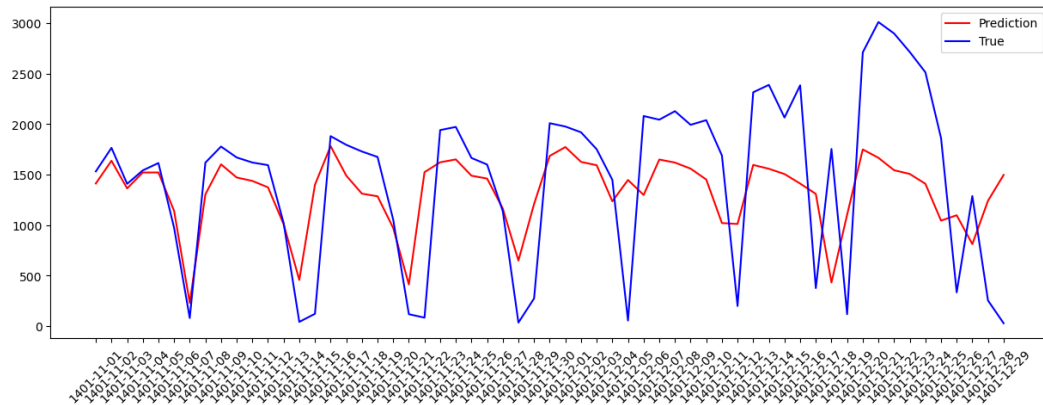


دیتاست ترکیبی

مقادیر واقعی و پیشبینی شده داده های validation با

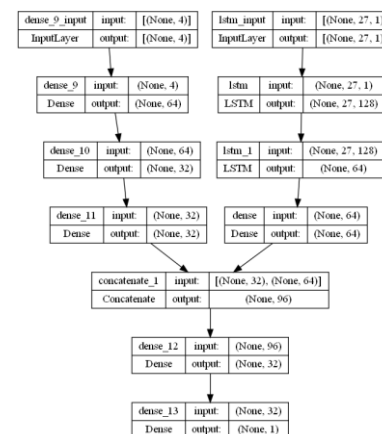
$$863/1 = \text{rmse}$$

شبکه عصبی LSTM



دیتاست نوع ۲

داده های validation
با $rmse = 70.1/4$



دیتاست ترکیبی

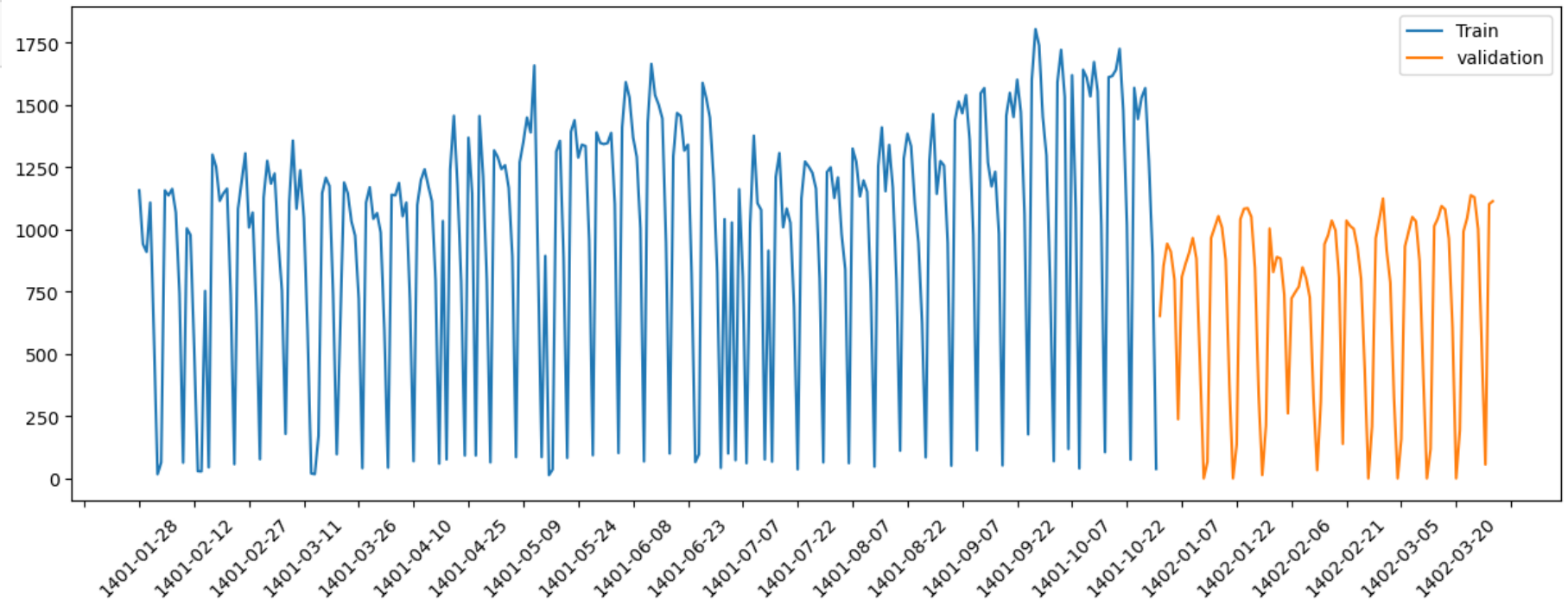
داده های validation
با $rmse = 63.0/5$

پیشبینی بهار ۱۴۰۲

m2_test_df['pred']

1402-01-01	652.562927
1402-01-02	853.857544
1402-01-03	942.809387
1402-01-04	910.538879
1402-01-05	797.508362
...	
1402-03-27	522.794067
1402-03-28	55.489887
1402-03-29	1101.873901
1402-03-30	1113.702148
1402-03-31	NaN

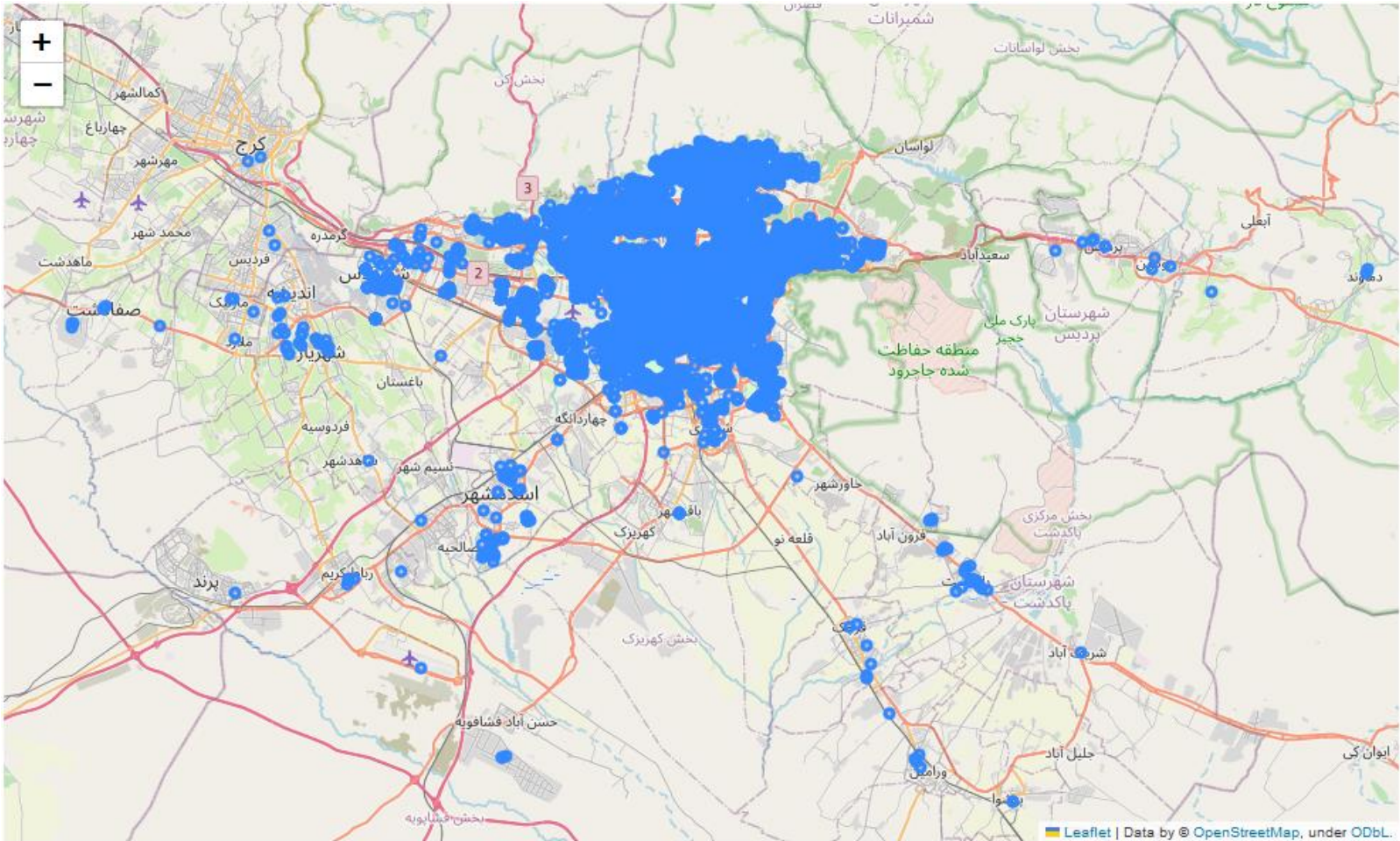
Name: pred, Length: 93, dtype: float64



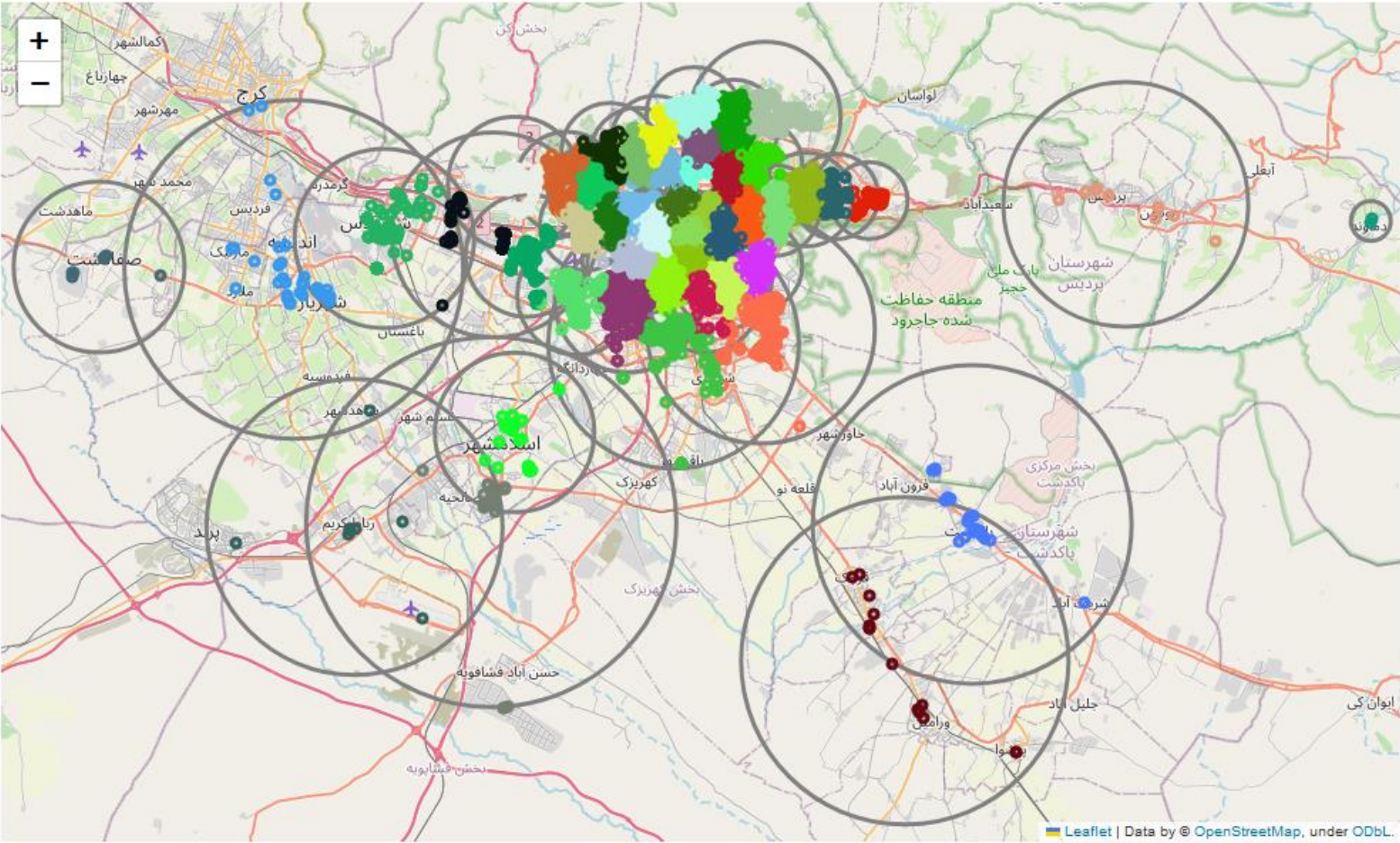
پیشبینی بهار ۱۴۰۲

سوال ۲:

■ نقشه قبل از خوشه بندی



سوال ۲:



■ نقشه بعد از خوشه بندی

سوال ۲:

جدول نتایج

شماره سفارش	Lats	Longs	cluster
10118716236728	51.408035	35.695091	24
10118716236689	51.408035	35.695091	24
10118716254610	51.450652	35.803117	46
10118716254504	51.274695	35.708660	7
10118716254509	51.418229	35.761035	16
...
10118716973840	51.358150	35.714482	0
10118716978840	51.524781	35.735659	37
10118716977546	51.417481	35.718140	2
10118716978011	51.406347	35.707970	2
10118716978660	51.411183	35.720230	2

23339 rows × 3 columns