



چالش مرحله دوم

یافتن پرسودترین برنامه‌ی پروازی

هدف مرحله‌ی اول مسابقه آشنایی شرکت‌کننده‌ها با داده‌های مسئله، پیش‌بینی تعداد تقاضا بر حسب تاریخ خرید بلیط و در نهایت برگزیدن تیم‌های برتر بود. در مرحله‌ی دوم بنا داریم که چالش کاربردی دیگری را که به نحوی به چالش مرحله‌ی اول نیز ارتباط پیدا می‌کند به عنوان چالش اصلی مسابقه مطرح کنیم. به طور کلی هدف از هرگونه پیش‌بینی از جمله پیش‌بینی تقاضای خرید بلیط، تنظیم «برنامه‌ی پروازی» ایرلاین‌ها برای داشتن بیشترین بازدهی است. با توجه به این موضوع، چالش این مرحله از مسابقه داده کاوی امیرکبیر، «ارائه‌ی برنامه‌ی پروازی هواپیما» در نظر گرفته شده است.

سناریو

در این مرحله نیز از داده‌های خرید بلیط سال‌های ۱۳۹۵ و ۱۳۹۶ ارائه شده در مرحله‌ی اول مسابقه استفاده می‌شود. با استفاده از این داده‌ها نیاز به طراحی مدلی است که برنامه‌ی پروازی یک هواپیما که در ابتدای روز شهر ۷۰ قرار دارد را برای تاریخ‌های مشخصی ارائه نماید. در نهایت و برای ارزیابی برنامه‌ی پروازی ارائه شده توسط هر تیم امتیاز هر برنامه‌ی پروازی متناسب با تعداد مسافران واقعی آن پرواز و همچنین قیمت بلیط پرواز محاسبه خواهد شد. (نحوه‌ی محاسبه امتیاز به طور دقیق‌تر در بخش‌های بعدی توضیح داده می‌شود.) با توجه به نحوه‌ی محاسبه‌ی امتیاز، شما برای تنظیم برنامه‌ی پروازی لازم است تا تعداد مسافران هر پرواز را پیش‌بینی کنید و با توجه به تعداد مسافرهای پیش‌بینی شده و همچنین قیمت پرواز هر مسیر، پرسودترین ترکیب پروازها را برای یک هواپیما ارائه نمایید. دقت شود که در این مرحله میزان تقاضای تاریخ و زمان **پرواز** مد نظر است (متفاوت از زمان خرید بلیط که در مرحله اول مطرح بود).



فرض‌ها و محدودیت‌ها

۱. اولین پرواز هر روز باید از مبدا شهر ۷۰ باشد.
۲. آخرین پرواز هر روز باید به مقصد شهر ۷۰ باشد.
۳. دنباله‌ی پروازها باید به هم مرتبط باشند. به عبارتی نمی‌توان از شهر ۷۰ به شهر ۳۰ رفته و در پرواز بعد در حالی که در شهر ۳۰ قرار داریم، از شهر ۴۰ به شهر ۴۹ رفت!
۴. فاصله دو پرواز برابر با ۳ ساعت در نظر گرفته می‌شود. به این معنی که مثلاً وقتی یک پرواز ساعت ۹:۱۵ دقیقه باشد، پرواز بعدی هواپیما می‌تواند از ساعت ۱۲:۱۵ به بعد صورت گیرد.
۵. هر پرواز فقط در دقیقه‌های بخش‌پذیر بر ۱۵ می‌تواند شروع شود (برای مثال ۸:۰۰، ۸:۱۵، ۸:۳۰ و ...).
۶. فقط یک هواپیما با ظرفیت ۲۰۰ نفر برای هر روز موجود است.
۷. در طی مسیرها تکرار شهرها مجاز است.

داده‌ها

داده‌های ارائه شده مربوط به خریدهای بلیط از سایت علی‌بابا است که به همراه صورت مسئله در سایت بارگذاری شده است. در ادامه چند نکته در مورد فایل‌های ضمیمه شده آورده شده است:

۱. هر سطر از داده‌ها نشانگر یک درخواست خرید بلیط است.
۲. خط اول هر فایل داده حاوی نام ویژگی‌ها است.
۳. در تمام سطرها، علامت "،" ستون‌ها را از هم جدا می‌کند.
۴. علامت "." نشانگر جدا کننده اعشار است.

فایل "AUT DMC 2018 - Features.pdf" موجود در "task2.zip" حاوی لیستی از تمام ویژگی‌ها و توضیحات مربوط به آن ویژگی‌ها است.

دقت کنید که امکان وجود داده‌ی پرت وجود دارد و تیم‌ها بنا به صلاحدید خود می‌توانند هر پردازشی روی آن‌ها انجام دهند.



بارگذاری فایل پیش‌بینی

تیم‌ها بایستی پروازهای پیشنهادی برای روزهای ۴، ۱۲، ۱۴، ۱۸، ۲۰، ۲۴ و ۲۹ شهریور و ۵، ۱۲، ۱۵، ۱۸، ۲۱، ۲۵ و ۲۷ مهر سال ۱۳۹۷ را ارائه دهند. هر تیم در هر روز می‌تواند حداکثر ۶ فایل پیش‌بینی بارگذاری کرده و امتیاز خود را در کنار امتیاز تیم‌های دیگر مشاهده کند. امکان بارگذاری نتایج تا ۹ اسفند وجود خواهد داشت.

نام ستون	توضیحات	محدوده‌ی مقادیر
Date	تاریخ پرواز	YYYY/MM/DD
Time	ساعت و دقیقه پرواز	HH:MM
From	مبدا مسیر	اعداد صحیح مثبت
To	مقصد مسیر	اعداد صحیح مثبت

فایل نهایی باید دقیقاً با فرمت بالا باشد. مثال زیر یک بارگذاری صحیح را نشان می‌دهد:

Date,Time,From,To

1397/06/04,7:15,70,20

1397/06/04,11:30,20,49

1397/06/04,15:45,49,70

...

برای فایل نهایی از فرمت CSV استفاده شود.

معیار ارزیابی

پس از ارائه‌ی برنامه‌ی پروازی برای روزهای خواسته شده، میزان سود حاصل برای هر پرواز f به صورت رابطه‌ی زیر محاسبه می‌شود. سود نهایی برابر با حاصل جمع سود پروازهای پیشنهاد شده خواهد بود.

$$Profit_f = \min(N, 200) * Price$$



N نشان دهنده تعداد مسافر واقعی بالقوه برای مسیر پرواز در همسایگی یک ساعته مربوط به پرواز است. برای مثال اگر شما در برنامه‌ای که ارائه می‌دهید، پروازی در ۴ شهریور ساعت ۸:۱۵ از شهر ۷۰ به ۴۰ در نظر گرفته‌اید، تعداد مسافرانی که در ۴ شهریور ۹۷ در بازه‌ی زمانی ۷:۱۵ تا ۹:۱۵ از شهر ۷۰ به ۴۰ رفته‌اند به عنوان تعداد مسافر واقعی بالقوه (N) در نظر گرفته می‌شود.

$Price$ نشان دهنده میانگین قیمت سال پیش بازه‌ی زمانی متناظر مسیر پرواز است. قیمت پروازها به صورت میانگین قیمت همان روز در سال قبل در ۴ بازه‌ی زمانی زیر برای هر مسیر در نظر گرفته می‌شوند.

۰۵:۵۹ تا ۰۰:۰۰	۰۶:۰۰ تا ۱۱:۵۹	۱۲:۰۰ تا ۱۷:۵۹	۱۸:۰۰ تا ۲۳:۵۹
----------------	----------------	----------------	----------------

مثلاً قیمت پروازی که ۴ شهریور ۹۷ ساعت ۸:۰۰ از شهر ۷۰ به ۴۰ حرکت می‌کند، برابر با میانگین قیمت تمام پروازهای روز مشابه سال ۹۶ در بازه‌ی زمانی ۰۶:۰۰ تا ۱۱:۵۹ بین شهرهای ۷۰ و ۴۰ فرض می‌شود. دقت شود که منظور از تاریخ پرواز `Departure_Date` است.

گزارش و کد

هر تیم باید اقدام به تهیه‌ی گزارشی در مورد روش حل مسئله و مدل انتخابی و همچنین دید بیزینسی مسئله کند. توجه شود که بررسی گزارش‌ها توسط تیم علی‌بابا صورت می‌گیرد و دید بیزینسی از اهمیت بیشتری برخوردار است. همچنین کد استفاده شده برای تولید نتایج نیز باید بارگذاری شود تا امکان بررسی صحت کار و جلوگیری از تقلب وجود داشته باشد. امکان بارگذاری گزارش و کد تا **۹ اسفند** وجود خواهد داشت.

ارزیابی نهایی

امتیاز نهایی تیم‌ها با در نظر گرفتن ضریب ۸۰٪ برای معیار ارزیابی و ۲۰٪ برای گزارش نهایی محاسبه خواهد شد. برای این منظور امتیازهای پرت حذف شده و تیم‌ها با توجه به امتیاز کسب شده از ۰ تا ۸۰ نمره خواهند گرفت که با نمره‌ی گزارش جمع شده و نمره‌ی نهایی به دست خواهد آمد. لازم به تاکید مجدد است که امتیاز گزارش، بیش از هر چیز از «نگاه شما به مساله»، «روش فکر» و به طور کلی از وجود دید بیزینسی در گزارش حاصل می‌شود.