

پروژه درس مبانی برنامه سازی

تحلیل اطلاعات پرواز

استاد: آقای علیرضا کدیور

راهنما: آقای علی غلامزاده

اعضای گروه:

زهره سادات بحری

امیرمهدی دهقان روزی

حمیدرضا نوری

مهندسی صنایع

دانشگاه صنعتی شریف

نیم سال اول ۹۸-۹۹

فهرست مطالب

| | |
|------------------------------------|----|
| مقدمه | ۳ |
| گزارش تقسیم کار | ۴ |
| توصیف داده | ۵ |
| قابلیت به روزرسانی | ۶ |
| توضیح کدهای پایتون | ۷ |
| توضیح کد Quality | ۷ |
| توضیح کد CleanFileExporter | ۷ |
| توضیح کد E part | ۷ |
| توضیح کد سوال ۱ | ۷ |
| توضیح کد سوال ۲ | ۷ |
| توضیح کدهای سوال ۳ | ۸ |
| توضیح کد سوال ۴ | ۸ |
| توضیح کد سوال ۵ | ۸ |
| توضیح کد سوال ۶ | ۸ |
| توضیح کد سوال ۷ | ۸ |
| گزارش توصیفی از اطلاعات تجمیع کلان | ۱۰ |
| کیفیت سنجی | ۱۱ |
| گزارش تحلیلی و پاسخ به پرسش‌ها | ۱۳ |
| پرسش ۱ | ۱۳ |
| پرسش ۲ | ۱۴ |
| پرسش ۳ | ۱۵ |
| پرسش ۵ | ۲۰ |
| پرسش ۶ | ۲۳ |
| پرسش ۷ (پرسش جدید مطرح شده) | ۲۹ |

مقدمه

حتما شما نیز با هواپیما سفر کرده‌اید. اگر بخواهید همین لحظه اقدام به خرید یک بلیت هواپیما بکنید با انتخاب‌ها و نگرانی‌هایی مواجه می‌شوید که برای همه‌ی ما آشنا است. دغدغه‌هایی از جمله: هزینه‌ی بلیت چقدر خواهد شد؟ بلیت کدام ایرلاین را تهیه کنم تا از کیفیت هواپیما و سفر اطمینان داشته باشم؟ کیفیت بالاتر معمولا با قیمت بیشتر همراه است، چقدر حاضرم پول بیشتری برای شرایط بهتر بپردازم؟

همچنین اگر در تعطیلات با هواپیما سفر کرده باشید، قطعاً به این فکر کرده‌اید که چندماه زودتر بلیت خود را تهیه کنید چرا که بخاطر آمار زیاد سفر در تعطیلات ممکن است ظرفیت پروازها تکمیل شود و البته... شما هم حس می‌کنید سرویس‌های ارائه دهنده در زمان تعطیلات نوروز هزینه‌های خود را بالاتر می‌برند؟

در این پروژه پاسخ این پرسش‌های معمول که برای همه‌ی ما پیش آمده و همچنین پرسش‌های دیگری را با استفاده از اطلاعات پرواز یک منبع، به صورت کمی و ریاضیاتی به دست می‌آوریم. داده‌های ما شامل داده‌هایی از جمله شناسه کاربری مشتریان، تاریخ و ساعت خرید، تاریخ و ساعت حرکت بلیت و قیمت آن، از ابتدای سال ۹۵ تا پایان سال ۹۶ است. این گزارش شامل پاسخ و تحلیل پرسش‌های مطرح شده در پروژه، توضیح کدهای ارائه شده همراه با آن و توصیف و توضیح کمی و کیفی اطلاعات پرواز است. برای سهولت در درک مطلب خروجی تمام کدها به صورت نمودار و یا جدول است که تصویر تمام آن‌ها در این گزارش قرار داده شده است. راستی! اگر داده‌ی مشابهی دارید که می‌خواهید نتیجه‌ی پرسش‌ها و نمودارها را برای آن ببینید می‌توانید به راحتی این کار را انجام دهید، در ادامه نحوه‌ی انجام این کار را به شما توضیح داده‌ایم. امیدواریم بعد از خواندن این گزارش اطلاعات خوبی در مورد عملکرد اشخاص و ایرلاین‌ها در خصوص سفرهای هوایی به دست آورید.

گزارش تقسیم کار

قبل از پاسخ به پرسش‌ها باید کیفیت سنجی داده‌ها انجام شود و فایل تمیز قابل استفاده‌ای از اکسل دیتاهای در دست ساخته شود. با توجه به سختی سوالات، ارتباط سوالات با یکدیگر و تعداد نفرات پرسش‌ها را تقسیم می‌کنیم. با تحلیل جزئی سوالات می‌فهمیم که سوال ۱ با ۲ (هر دو در مورد میزان خرید مشتری‌ها) و سوال ۳ با ۴ (هر دو مرتبط با زمان خرید بلیت‌ها) مرتبط هستند و سوالات ۵ و ۶ باقی می‌مانند. هر یک از اعضای گروه یک جفت از سوالات را انتخاب کرد. سپس یکی از اعضای گروه به عنوان داوطلبانه طرح پرسش جدید را به عهده گرفت. در آخر تصمیم گرفتیم قسمت مقدمه، گزارش توصیفی از اطلاعات تجمیع کلان، کشیدن دیاگرام به روز رسانی، توصیف داده و گزارش تقسیم کار نیز بعداً با توجه به زمان اتمام پرسش‌ها به اعضای گروه محول شود.

تقسیم کار انجام شده

زهرا سادات بحری: پرسش‌های ۵ و ۶ و پرسش اختیاری، کیفیت سنجی و گزارش توصیفی از تجمیع کلان و طراحی دیاگرام بخش به روز رسانی

امیرمهدی دهقان روزی: پرسش‌های ۳ و ۴، نوشتن مقدمه و گزارش استراتژی و تقسیم کار

حمیدرضا نوری: پرسش‌های ۱ و ۲، توصیف داده

توصیف داده

| id | user_id | request_date_id | request_time | departure_date_id | departure_time_id | company | source | destination | price | month | path |
|--------|---------|-----------------|--------------|-------------------|-------------------|---------|--------|-------------|-------------|-------|-------|
| 100015 | 218660 | 20160321 | 20:19:00 | 20160322 | 2015 | 9 | 30 | 42 | 437.610619 | 3 | 30-42 |
| 100016 | 85 | 20160321 | 20:32:23 | 20160322 | 2015 | 9 | 30 | 42 | 437.610619 | 3 | 30-42 |
| 100017 | 32275 | 20160321 | 21:22:00 | 20160323 | 1245 | 18 | 66 | 37 | 313.495575 | 3 | 66-37 |
| 100018 | 296909 | 20160321 | 22:54:00 | 20160401 | 750 | 3 | 15 | 58 | 2715.707964 | 4 | 15-58 |
| 100019 | 296909 | 20160321 | 22:54:00 | 20160401 | 750 | 3 | 15 | 58 | 2715.707964 | 4 | 15-58 |

شکل ۱ - چند سطر نمونه از فایل آماده سازی شده

ستون ۱ (id): هر id مربوط به یک بار خرید بلیط است و با خرید بلیط های بعدی یک id جدید خواهیم داشت. ستون های بعدی در واقع اطلاعاتی مربوط به این ستون به ما می دهد.

ستون ۲ (user_id): این ستون مربوط به شخص خریدار بلیط بوده و قابلیت تکرار شدن را دارد. برای مثال اگر کسی دوبار بلیط تهیه کرده باشد user_id شخص دوبار تکرار خواهد شد.

ستون ۳ (request_date_id): این بخش مربوط به تاریخ خرید بلیط است یعنی روزی که شخص بلیط را خریداری کرده.

ستون ۴ (request_time): این ستون ساعت خرید بلیط را به ما نشان می دهد.

ستون ۵ (departure_date_id): در این ستون تاریخ حرکت هواپیما به سمت مقصد تعیین شده.

ستون ۶ (departure_time_id): در این ستون ساعت حرکت هواپیما به طرف مقصد مشخص شده.

ستون ۷ (company): در این ستون اعدادی را مشاهده می کنیم که هر عدد مربوط به یک شرکت هواپیمایی خاص است. درواقع به ما نشان می دهد که هر بلیط مربوط به کدام شرکت هواپیمایی است.

ستون ۸ (source): در این بخش مبدا های مختلف در پرواز با اعداد خاصی نشان داده شده اند و هر عدد مربوط به یک مبدا است.

ستون ۹ (destination): در این بخش هر عدد نشانگر یک مقصد خاص است و مقصد های مختلف با اعداد شماره گذاری شده اند.

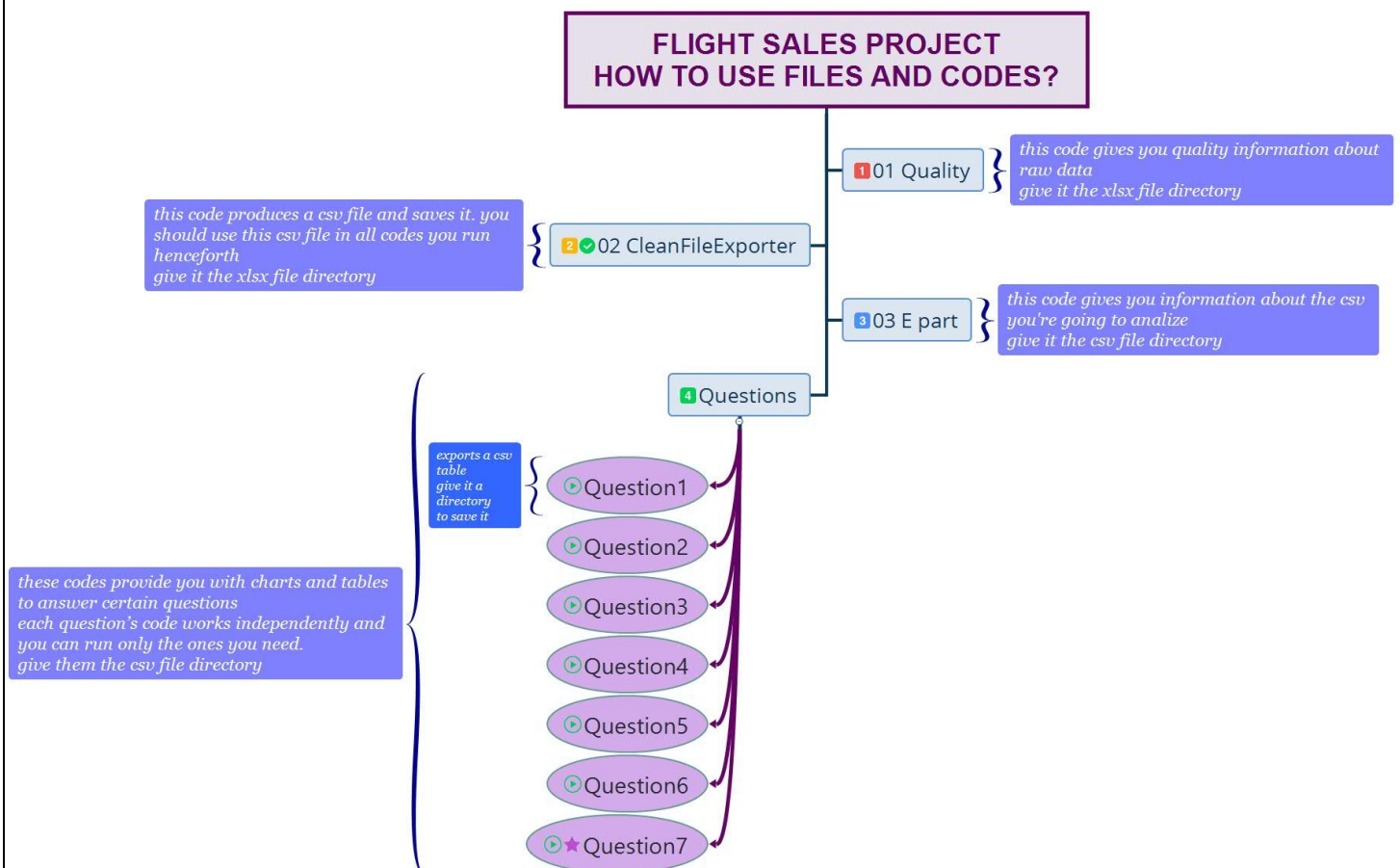
ستون ۱۰ (price): آخرین ستون هم قیمت هر بلیط را به ما نشان می دهد.

ستون ۱۱ (month): ماه حرکت هواپیما برای هر سفارش است که برای سهولت در پاسخ به پرسش ها اضافه شده است.

ستون ۱۲ (path): زوج مرتب مسیر (مقصد-مبدا) برای هر سفارش است که با ترکیب ستون ۸ و ۹ به دست آمده است.

قابلیت به روزرسانی

برنامه‌نویسی کدهای پروژه به این صورت است که اگر داده‌های قبلی به روز رسانی شوند یا جدول دیگری با ستون‌های مشابه (نام مشابه) داشته باشیم، می‌توانیم نمودارها، اعداد و جداول را برای آن به دست آوریم. کفایت داده‌های خود را با فرمت xlsx ذخیره کنید و آدرس آن را در ابتدای کدهای داده شده کپی کنید. سپس طبق دیاگرام زیر برای رسیدن به پاسخ پرسش‌های خود عمل کنید.



توضیح کدهای پایتون

توضیح کد Quality: این کد برای پاسخ به قسمت کیفیت سنجی است. فایل `xlsx` اصلی را به عنوان ورودی می‌گیرد. طول آن را چاپ می‌کند سپس برای شمارش تعداد داده‌های تکراری، ستون `id` را حذف می‌کند. تعداد سطرهای تکراری و تعداد کل داده‌های یکتا را پرینت می‌کند. سپس تعداد کل داده‌های خالی در هر ستون را پرینت می‌کند و سپس تعداد کل سلول‌های خالی در فایل را پرینت می‌کند. سپس سطر و ستون `NaN` ها و کل سطرهای حاوی داده‌ی خالی را به کاربر نشان می‌دهد. در مرحله بعد تعداد آیدی‌های نامعتبر (صفر و منفی) و مقادیر آنها را چاپ می‌کند. سپس تعداد صفرها در هر ستون و سپس تعداد صفرها در کل داده را چاپ می‌کند. نهایتاً می‌گوید چند درصد دیتا قابا تحلیل می‌باشد. همچنین هیستوگرام توزیع مقادیر ستون `price` را رسم می‌کند. سلول‌های این کد نباید بیش از یکبار `run` شوند در صورت اجرای دوباره به علت حذف شدن برخی سطرها، تعدادی از سلول‌ها با ارور مواجه خواهند شد.

توضیح کد CleanFileExporter: فایل `xlsx` اصلی را به عنوان ورودی می‌گیرد. دیتاهای غیرقابل استفاده را مانند کد `Quality` پیدا کرده و حذف می‌کند. سطرهای تکرار شده‌ای که تمام مقادیر ستون‌های آنها معتبر است را حذف نمی‌کند. چرا که این سطرها به معنای خرید چند بلیت همزمان توسط یک کاربر است. تعدادی ستون لازم برای پاسخ به سوالات (ستون ماه عزیمت و ستون مسیر) را به آن اضافه می‌کند. تایپ داده‌ها را `Int64` و برخی را `str` می‌کند (با توجه به تایپ لازم برای پاسخ به سوالات) و سپس فایل `CSV` جدیدی شامل تمام داده‌های قابل استفاده در پوشه محل ذخیره کد ذخیره می‌کند. از این فایل `CSV` برای پاسخ به پرسش‌ها استفاده می‌شود.

توضیح کد E part: این کد بسیار شبیه کد کیفیت سنجی است. فایل `CSV` را به عنوان ورودی می‌گیرد. تعداد سطر و ستون‌های آن و تایپ هر ستون و حجم کل فایل را چاپ می‌کند. سپس مقادیر یکتای هر ستون را چاپ می‌کند. نهایتاً مقادیر بیشینه و کمینه برای زمان و تاریخ درخواست، تاریخ حرکت، ایرلاین، بیشترین مبلغ و کمترین مبلغ و همچنین میانگین مبلغ را در خروجی چاپ می‌کند.

توضیح کد سوال ۱ - ابتدا فایل `FlightSalesClean.csv` را به عنوان ورودی می‌گیرد. سپس با استفاده از ستون `user_id` و `price` میانگین ستون `price` را برای هر کاربر بدست می‌آورد (حتی برای کاربر هایی که تعداد دفعات خریدشان بیش از یک بار است). این جدول را در قالب یک فایل `CSV` ذخیره می‌کند. سپس تعداد کاربران را محاسبه می‌کند. بعد داده‌های ستون `price` را جمع می‌کند. و در آخر مقدار پول خرج شده متوسط کل کاربران را بدست می‌آورد.

توضیح کد سوال ۲ - فایل `CSV` را به عنوان ورودی می‌گیرد. برای هر کاربر جمع کل هزینه را حساب می‌کند. سپس امتیاز هر کاربر را حساب می‌کند و در یک لیست می‌ریزد. محاسبه امتیاز به این صورت است که برای خریدهای بالای ۹۰۰ تومان، هر ۱۰۰ تومان خرید بیشتر ۵ امتیاز به کاربر اضافه می‌کند. سپس

لیست امتیاز هارا به جدول اضافه کرده و امتیاز ها و شناسه کاربر و میزان کل خرید ۱۰ کاربر پرامتیاز را چاپ می‌کند.

توضیح کدهای سوال ۳- فایل csv را به عنوان ورودی میگیرد. با استفاده از دو ستون `departure_date_id` و `request_date_id` نمودار های تعداد درخواست و تعداد پرواز بر حسب ساعت های یک شبانه روز، روز های مختلف یک هفته و ماه های یک سال به عنوان خروجی ترسیم میکند. همچنین `min` و `max` نمودارهای ترسیم شده تعداد درخواست بر حسب ساعت، روز و ماه را به عنوان خروجی چاپ می‌کند.

توضیح کد سوال ۴- فایل csv را به عنوان ورودی میگیرد. با دو ستون `request_date_id` و `departure_date_id` نمودار اختلاف روز بین ساعت خرید بلیط و ساعت پرواز را به عنوان خروجی ترسیم میکند.

توضیح کد سوال ۵- فایل csv را به عنوان ورودی میگیرد. پرتکرارترین عناصر ستون های `source` ، `destination` و `path` را پیدا کرده، نمودار و جدول ۵ عنصر اول هر کدام به همراه تعداد تکرار آنها را به عنوان خروجی تحویل میدهد. نمودار پرتکرارترین مبداها، پرتکرارترین مقصدها، پرتکرارترین مسیر ها و همچنین جدول مربوط به هر نمودار را ترسیم میکند.

توضیح کد سوال ۶- فایل csv را به عنوان ورودی میگیرد. ابتدا برای کل داده‌ها، مبلغ کل هر کمپانی را محاسبه میکند. نمودار دایروی مربوط به درآمد کل کمپانی‌ها، نمودار میله‌ای ۵ کمپانی پردرآمد و جدول مربوط به آن را پرینت می‌کند. سپس تعداد تکرار هر کمپانی را محاسبه می‌کند و نمودار دایروی مربوط به تعداد تکرار کل کمپانی‌ها، نمودار میله‌ای ۵ کمپانی پرتکرار و جدول مربوط به آن را پرینت می‌کند. برای مقایسه‌ی بهتر نمودار میله‌ای تعداد کل و مبلغ کل ۵ کمپانی اول را با هم در یک نمودار ترسیم و پرینت میکند. سپس جمع زدن مبلغ و تعداد هر کمپانی را به صورت جداگانه برای هرماه و هرماه انجام می‌دهد. ۲ نمودار میله‌ای دسته‌ای رسم میکند که هر دسته متعلق به یک ماه است و در هر ماه میله‌های ۳ کمپانی برتر ترسیم می‌شود؛ یک نمودار برای پردرآمدترین های هرماه و یک نمودار برای طرفدارترین های هرماه. سپس اینکار را برای مبدا انجام می‌دهد و در خروجی نمودار های ۳ کمپانی برتر در هر مبدا را برای مبدهایی با بیشتر از ۵ کمپانی رسم می‌کند؛ یک نمودار برای پردرآمدترین های مبدهای مورد نظر و یک نمودار برای پرتطرفدارترین های مبدهای مورد نظر. همچنین نهایتا برای شهرهایی که از لیست حذف کرده بود به دلیل اینکه از تعداد ایرلاین های کافی استفاده نکرده بودند یک جدول شامل نام کمپانی و مجموع دفعات استفاده شده در این مبداها نمایش می‌دهد.

توضیح کد سوال ۷- فایل csv را به عنوان ورودی میگیرد. برای هرماه مبلغ کل بلیت‌های هواپیماهایی که در همان ماه حرکت می‌کنند را محاسبه کرده، نمودار دایروی و نمودار میله‌ای آن به همراه جدول ماه و مبلغ کل را پرینت می‌کند.

تمیز کردن و آماده سازی فایل

فایل خام xlsx موجود کاملاً برای تحلیل مناسب نیست. تعدادی از خانه‌های جدول خالی می‌باشد که گزارش آن‌ها بدین شرح است: ستون departure_time_id دارای ۲ مقدار خالی (NaN) ستون Source دارای ۶ مقدار خالی و ستون Destination دارای ۵ مقدار خالی است. مجموعاً ۱۳ سلول دارای مقدار NaN در این جدول وجود دارد. همچنین یک آیدی نامعتبر با مقدار -۱ وجود دارد. جدول داده‌ها مجموعاً دارای ۱۲۸۶۷ سلول با مقدار صفر وجود دارد که تمام آن‌ها در ستون departure_date_id هستند.

داده‌ی خام در دست مجموعاً شامل ۲۱۱۷۷۵ سطر است. در فایل csv، از بین این سطرها، تمام سطرهایی که حداقل یکی از ستون‌های آن مقدار NaN یا صفر یا آیدی نامعتبر (صفر و منفی) برای تحلیل پرسش‌ها حذف شده‌اند. این فایل csv که کد CleanFlieExporter آن را تولید و سیو می‌کند، فایلی است که برای کیفیت سنجی و توصیف داده و پاسخ به پرسش‌ها از آن استفاده می‌کنیم.

گزارش توصیفی از اطلاعات تجمیع کلان

داده‌ی آماده سازی شده مجموعاً شامل ۱۹۸۸۱۴ سطر است. که از این تعداد ۶۸۳۸۰ سطر آن تکرار شده و در نتیجه تعداد سطرهای یکتا به ۱۴۳۳۹۵ تقلیل می‌یابد. در این پروژه در نظر گرفتیم که سطرهای تکراری خرید چند بلیت همزمان توسط یک کاربر است و از آن‌ها در تحلیل سوالات استفاده کردیم.

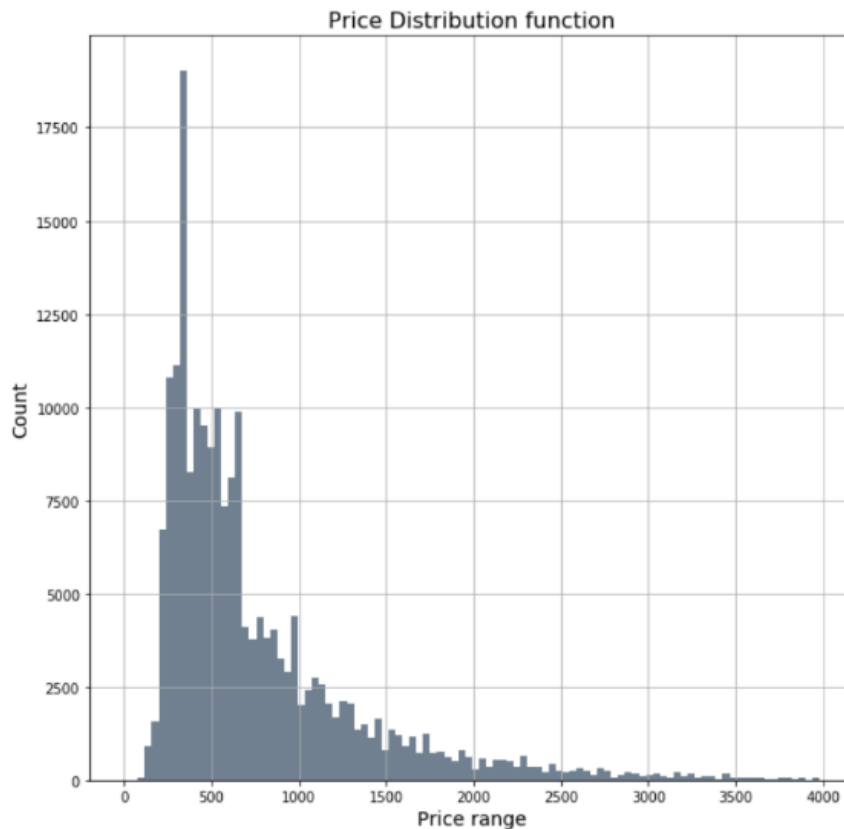
از آنجا که در فایل قابل استفاده تعداد کل سفارشات ۱۹۸۸۱۴ است و تعداد سطرهای دیتای خام برابر ۲۱۱۷۷۵ است، یعنی ۹۳,۸۸ درصد دیتای خام قابل تحلیل و استفاده است. که کل این داده‌های قابل تحلیل درباره‌ی ۱۳ ایرلاین، ۲۸ شهر به عنوان مبدا و مقصد، مجموع ۱۴۶ مسیر دوتایی مبدا و مقصد (زوج مرتب) و در نتیجه میانگین ۷۳ مسیر رفت و برگشت (یعنی به طور میانگین مسیر هر شهر با ۲ الی ۳ شهر دیگر) و مجموع ۴۱۸۱۱ کاربر و در ۱۲ ماه در سال‌های ۲۰۱۶ تا ۲۰۱۸ (نوروز ۹۵ تا ۹۷) است. از آنجایی که بازه‌ی سفارشات ۲ سال است و هر سال ۳۶۵ روز است، پس ما برای تمام روزهای این دوسال سفارش داریم چرا که مقادیر یکتای ستون request_date_id که نشان دهنده‌ی روزهای مختلف سفارشات است برابر ۷۳۰ است.

بازه‌ی دقیق تاریخ سفارشات از ساعت ۰۰:۰۰:۰۰ ۲۱ مارس ۲۰۱۶ تا ساعت ۰۰:۰۰:۰۰ ۲۰ مارس ۲۰۱۸ (دقیقاً دوسال) است. همچنین بازه‌ی ساعت ثبت سفارش‌ها (به ترتیب مینیمم و ماکسیمم این ستون) ساعت ۰۰:۰۰:۰۰ تا ساعت ۲۳:۵۹ است. بازه‌ی ماه‌ها نیز از ۱ تا ۱۲ است که البته با توجه به بند قبل که نتیجه گرفتیم در هر روز داده داریم نتیجه‌ای بدیهی است. مینیمم تاریخ و ساعت حرکت هواپیماها ساعت ۰۰:۰۰:۰۰ ۲۱ مارس ۲۰۱۶ و ماکسیمم آن ساعت ۰۰:۰۰:۰۰ ۱۵ آوریل ۲۰۱۸ می‌باشد.

در ستون قیمت می‌بینیم که کمترین مبلغ ۳۰ هزار تومان و بیشترین مبلغ ۹ میلیون و ۴ هزار تومان و میانگین ۷۹۰ هزار تومان است. به نظر می‌رسد که قیمت‌ها از قیمت واقعی بیشتر باشد. چرا که این میانگین تقسیم مبلغ کل بر کل سفارشات و در نتیجه میانگین قیمت یک بلیت است. در حقیقت ۷۹۰ هزار تومان برای خرید یک بلیت مبلغ زیادی است. اما چون تمام داده‌ها به یک نسبت شیفت خورده و افزایش می‌یابد این مشکل در تحلیل اطلاعات به صورت نسبی و مقایسه‌ای بین ماه‌ها و کمپانی‌ها و مشتری‌ها ایرادی ایجاد نمی‌کند. در مورد اینکه چرا قیمت‌ها ممکن است بیشتر از

کیفیت سنجی

هیستوگرام مربوط به ستون قیمت‌ها برای بررسی وجود داده‌ی پرت نیز رسم شده است. علت رسم این تابع توزیع این است که مشاهده می‌کنیم کمترین مبلغ ۳۰ هزار تومان و بیشترین مبلغ ۹ میلیون و ۴ هزار تومان است که هر دو با توجه به میانگین مبلغ‌ها که ۷۹۰ هزار تومان است داده پرت به نظر می‌رسند.



اما همانطور که پیداست داده‌ها حول یک نقطه (نقطه‌ای نزدیک به میانگین) جمع شده و دارای قله‌اند و با دور شدن از میانگین تعداد داده‌های پرت کمتر می‌شود. پس استفاده از این داده‌ها مشکلی ندارد و خطای اندکی در تحلیل‌های ما ایجاد خواهد کرد چرا که داده‌های پرت بسیار بزرگ با داده‌های پرت بسیار کوچک خنثی می‌شوند. همچنین تجمع داده‌ها در سمت چپ نمودار است. بدیهی است که مردم ترجیح می‌دهند با بلیت با قیمت کمتر را خریداری کنند. اما قله روی مبالغ خیلی کم نمی‌افتد چون در اینجا کیفیت‌های مختلف با قیمت‌های مختلف را مقایسه می‌کنیم و افراد حاضرند تا یک حدی برای شرایط بهتر پول بهتری بپردازند. بعد از گذشتن از حداقل شرایط مناسب، با افزایش قیمت، افراد کمتری حاضر به پرداخت پول بیشتر برای کیفیت بهتر هستند. برای همین نمودار نزول دارد و نمودار بعد از قله پیش می‌رود چرا که نهایتاً شخصی وجود دارد که به دلیل شرایط مالی بهتر و یا ضروری بودن سفر، بلیت با قیمت بیشتر را تهیه می‌کند. شکل نمودار با تابع توزیع نرمال متفاوت است از این رو که دلیل حاضر شدن افراد برای خرید بلیت با قیمت‌های بالاتر یکسان نیست. همچنین در بازه از صفر به قبل وجود ندارد (مبلغ منفی نداریم) در حالی که دامنه تابع توزیع نرمال از منفی بی‌نهایت تا مثبت بی‌نهایت است.

به غیر از بیشتر بودن تمام قیمت‌ها (به یک نسبت) و غلط بودن داده‌ی سطرهایی که مقدار صفر یا خالی یا منفی داشتند، داده‌ی خام در دست درست به نظر می‌رسد و فایل تمیزسازی شده قابل تحلیل و استناد است.

با توجه به قابل استفاده بودن ۹۳ درصد دیتای خام، کیفیت دیتای خام را خوب برآورد می‌کنیم. تعداد داده‌های فیلتر شده برای برآورد و تحلیل پرسش‌های مطرح شده مناسب است و می‌تواند نتایج درستی در نمودارها به نمایش بگذارد.

گزارش تحلیلی و پاسخ به پرسش‌ها

پرسش ۱- با کمک دیتای در دسترس، محاسبه کنید که به طور متوسط هر مشتری چند تومان خرید کرده است.

با توجه به آیدی کاربران درمیابیم که اغلب آن‌ها چندین بار اقدام به خرید بلیط کرده‌اند. پس با در دست داشتن قیمت خرید بلیط کاربران میتوان میانگین پولی که برای خرید بلیط‌ها خرج کرده‌اند را بدست آورد. حال برای محاسبه متوسط پول خرید بلیط توسط کاربران کفایت میانگین پول خرج شده توسطشان را به تعداد کاربران تقسیم کرد که نتیجه تقریباً برابر است با ۳ میلیون و ۷۶۰ هزار تومان.

جدول کاربران و میانگین خرید هرکدام در پوشه‌ی این پروژه ذخیره شده است.

پرسش ۲ - معیاری از خرید تعیین کرده و برای هر مشتری امتیاز محاسبه کنید. وفادارترین مشتری‌ها چه کسانی هستند؟

| | user_id | total expenditure (*10 ³ T) | score |
|---|---------|--|-------|
| 0 | 32502 | 1281083.628 | 64005 |
| 1 | 34737 | 662742.920 | 33090 |
| 2 | 5111 | 606659.491 | 30285 |
| 3 | 47441 | 557700.442 | 27840 |
| 4 | 55483 | 450809.734 | 22495 |
| 5 | 150606 | 435052.876 | 21705 |
| 6 | 155943 | 418113.274 | 20860 |
| 7 | 143356 | 373007.522 | 18605 |
| 8 | 37428 | 370335.398 | 18470 |
| 9 | 33865 | 361613.053 | 18035 |

روند امتیازدهی به این صورت است:

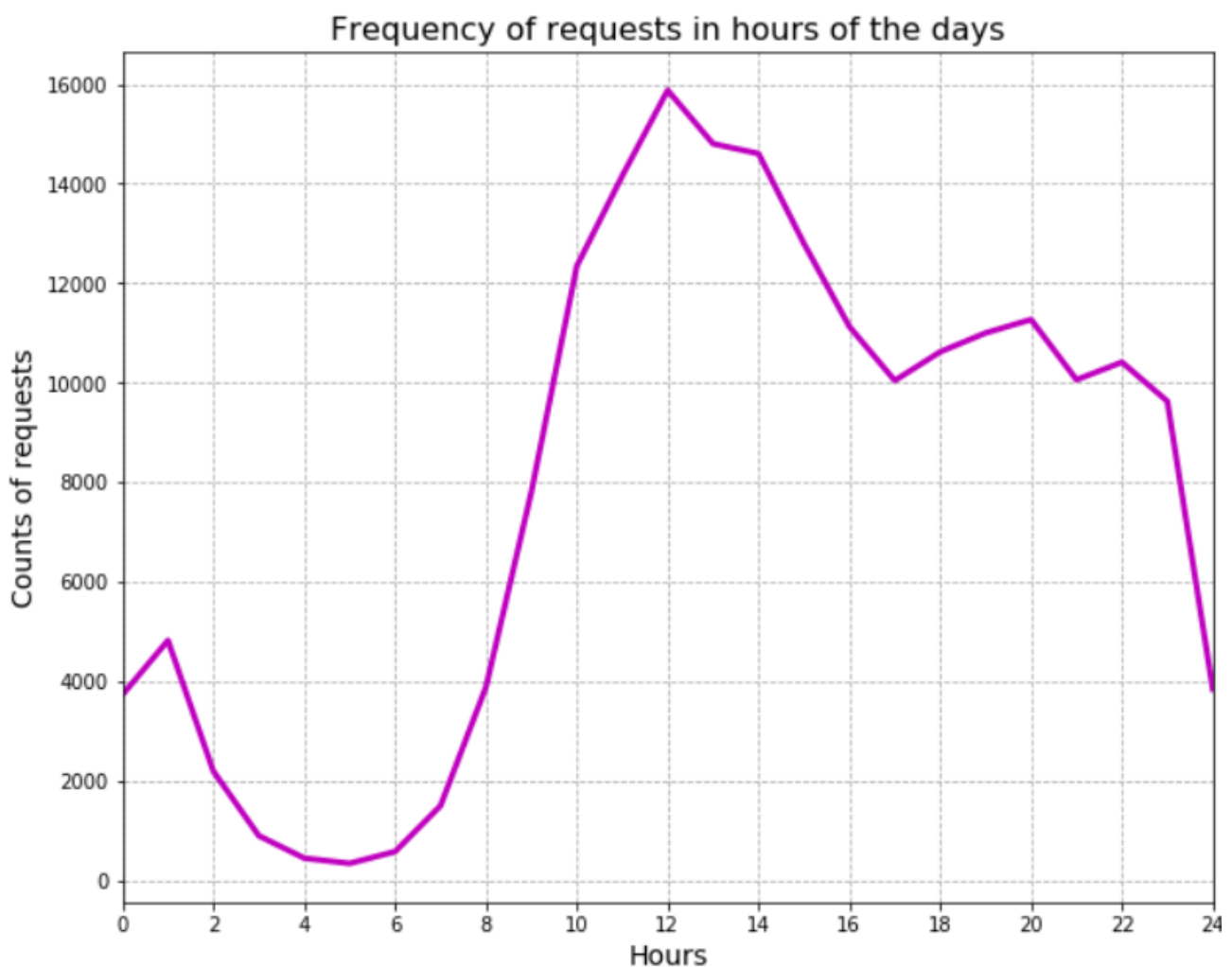
- ۱- به مشتریان با مجموع خرید کمتر از ۹۰۰ هزار تومان هیچ امتیازی تعلق نمی‌گیرد.
- ۲- مشتریانی که خرید کل آن‌ها بیشتر از ۹۰۰ هزار تومان است، با هر ۱۰۰ هزار تومان خرید بیشتر ۵ امتیاز دریافت می‌کنند. یعنی مشتری‌ای که مجموعاً ۹۵۰ هزار تومان خریده امتیازی نمی‌گیرد چرا که هنوز ۱۰۰ تومان بیشتر خرید نکرده. مشتری‌ای که ۱ میلیون تومان خرید کرده، ۵ امتیاز، یک میلیون و ۵۰ هزار تومان، ۵ امتیاز، یک میلیون و صد هزار تومان ۱۰ امتیاز دریافت می‌کند. فرمول محاسبه به این صورت است:

$$\left\lfloor \frac{x - 900}{100} \right\rfloor * 5$$

در جدول بالا نام ۱۰ کاربر با بالاترین امتیاز آورده شده است. از آنجایی که ماکسیمم مقدار ستون هزینه (یک خرید یک کاربر) ۹۴۰۰ بوده که معادل امتیاز ۴۲۵ است، پس مشتری‌هایی که در بالا با امتیازهایی از مرتبه‌ی ۱۰۰۰۰ آورده شده‌اند باید تعداد خریده‌های بسیار زیادی داشته باشند پس جزو مشتری‌های وفادار هستند.

پرسش ۳- وضعیت فروش را در روزهای مختلف هفته و ساعت های مختلف یک روز بررسی کنید. پرتعدادترین و کم پرتعدادترین زمان ها و روزهای هفته برای فروش کدام هستند؟ آیا مقدار فروش ماهیانه نسبت به ماه حساسیت دارد؟ (برای این کار تاریخ میلادی را به روز تبدیل می کنید).

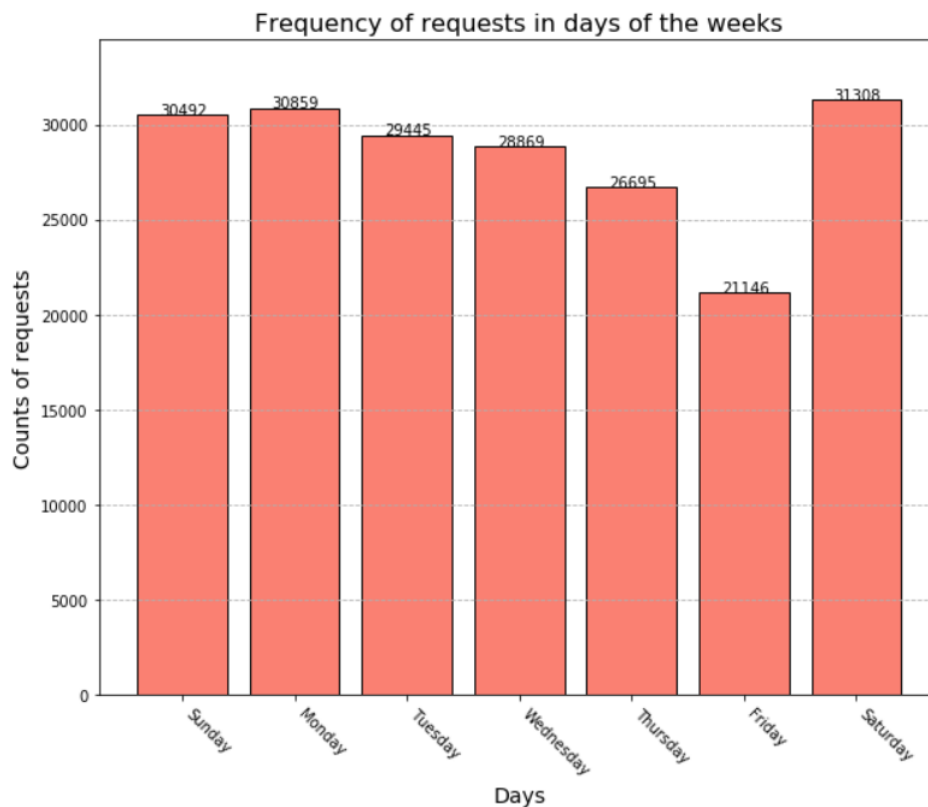
باتوجه به نمودار ۱-۳ که وضعیت فروش را در ساعت های مختلف یک روز نشان میدهد، مشخص می شود که بیشترین تعداد درخواست پرواز حول ساعت ۱۲ ظهر و کمترین تعداد درخواست حول ساعت ۵ صبح است. مشاهده می شود که از ساعت ۵ تا ساعت ۱۲ روند صعودی دارد و در ضمن ساعات اداری و زمان فعالیت مردم نیز بین این ساعات قرار دارند. همچنین دلیل روند نزولی نمودار بین ساعات ۱۲ تا ۱۶ احتمالا به دلیل استراحت نیم روزی مردم در این زمان است که بعد از آن کمی روند صعودی دارد و بعد ساعت ۸ بعد از ظهر بدلیل زمان خواب تقریبا روند نزولی کامل دارد.



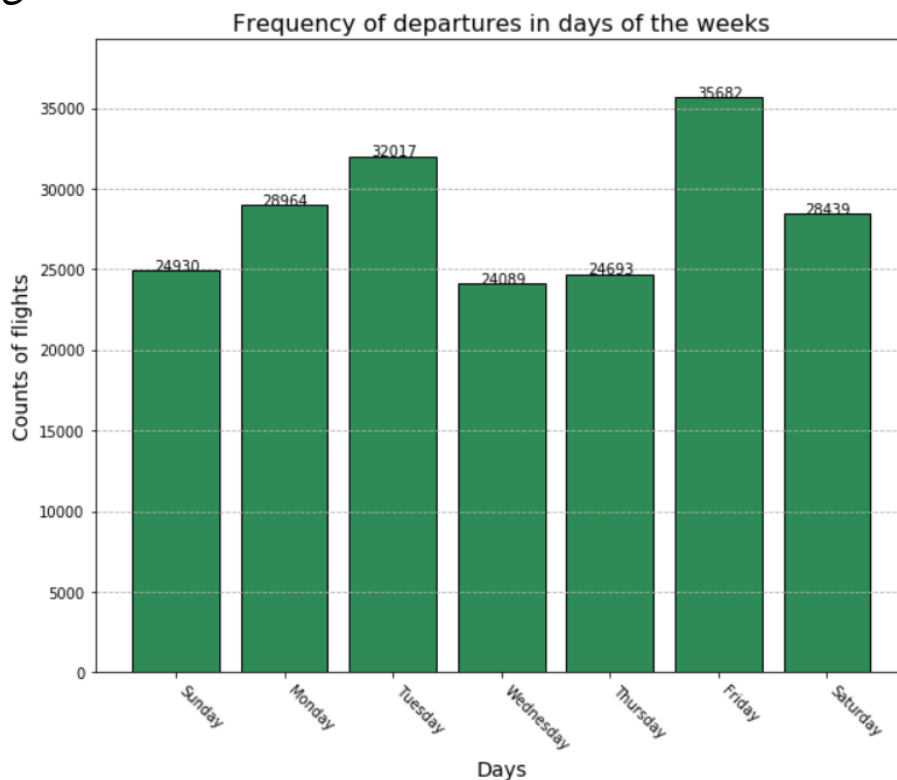
نمودار ۱-۳

پروژهی تحلیل اطلاعات پرواز | ۱۶

حال باتوجه به نمودار ۲-۳ که تعداد درخواست های پرواز در این سال ها را بر حسب روز های مختلف هفته نشان میدهد نتیجه می شود که روز Saturday با بیشترین تعداد درخواست پرواز محبوب ترین روز ازین نظر است. همچنین روز Friday با کمترین تعداد درخواست پرواز کم طرفدارترین است. پرتعدادترین بودن روز Saturday احتمالا بخاطر آخر هفته بودن و بهترین تایم برای درخواست پرواز برای آخر هفته جاری اش است.(یعنی خرید بلیت ابتدای هفته برای سفر آخر هفته) حتی باتوجه به نمودار ۳-۳، که نشان دهنده تعداد پرواز در روزهای مختلف هفته است، احتمال میرود بیشترین تعداد پرواز در روز Friday بدلیل اینکه تنها روز تعطیل اداری در طول هفته است، انجام شده است.(کشیدن نمودار تعداد پروازها در روزهای هفته و ماه جزو خواسته های پرسش نبودند و صرفا برای تحلیل بهتر کشیده شده اند. از این دو نمودار در تحلیل پرسش ۷ - پرسش طرح شده توسط اعضا- نیز استفاده شده است.)

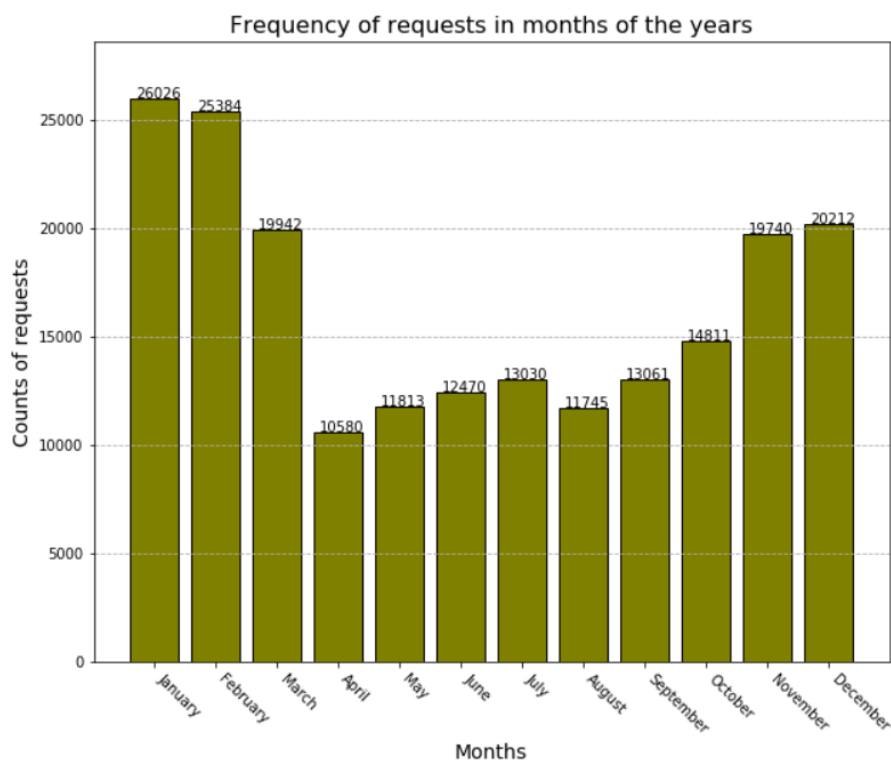


نمودار ۲-۳

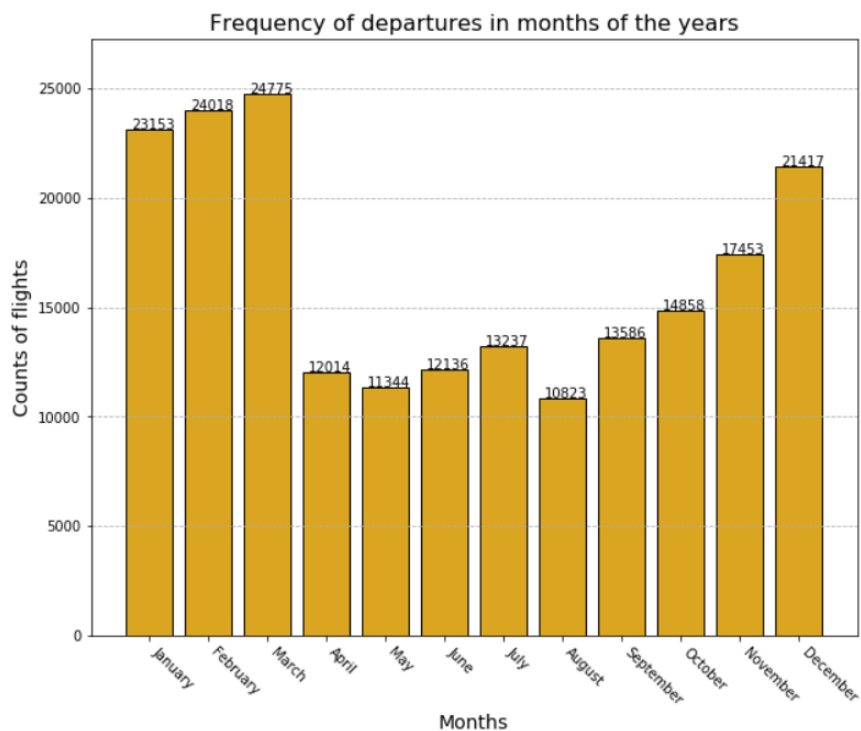


نمودار ۳-۳

نمودار ۳-۴ تعداد درخواست برای پرواز در هر ماه است. می‌بینیم که بیشترین تعداد درخواست پرواز مربوط به ماه January است و کمترین تعداد مربوط به ماه April. نتیجه این است که میزان خرید بلیط در ماه‌های سال میزان قابل توجهی تغییر می‌کند و این امر می‌تواند به دلیل وجود تعطیلات ثابت در هرماه، فصل امتحانات و مناسبت‌هایی از این دست باشد.



نمودار ۳-۴

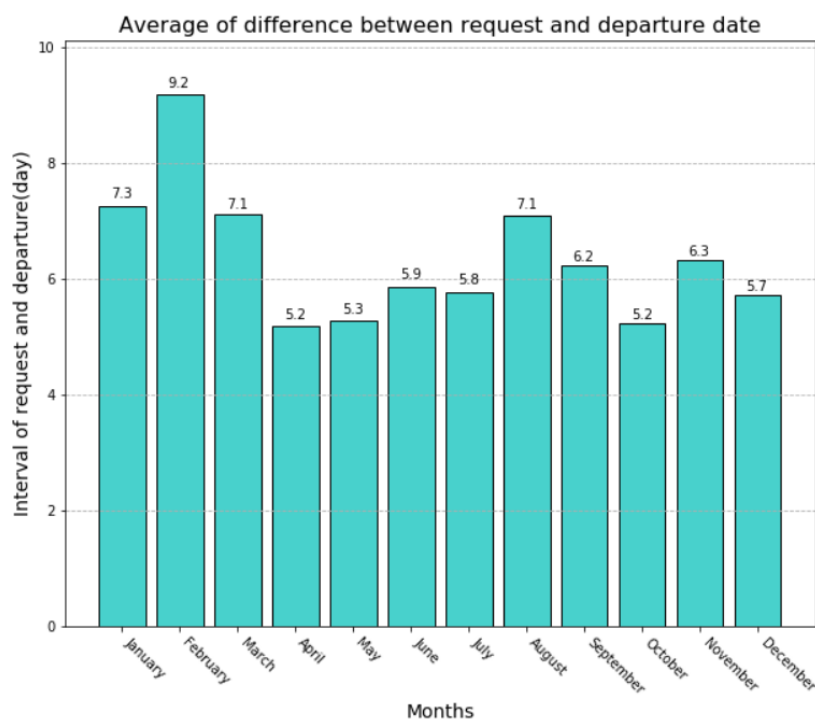


نمودار ۳-۵

حال نگاهی می اندازیم به نمودار ۳-۵ که نشان دهنده تعداد پرواز ها در ماه های مختلف یک سال است. اگر با دقت به دو نمودار نگاه کنیم می فهمیم که تعداد پرواز هر ماه متناسب با تعداد درخواست پرواز هر ماه است. در پرسش ۴ در مورد این موضوع بیشتر صحبت می کنیم. همچنین این نمودار را در پرسش ۷ دوباره تحلیل خواهیم کرد.

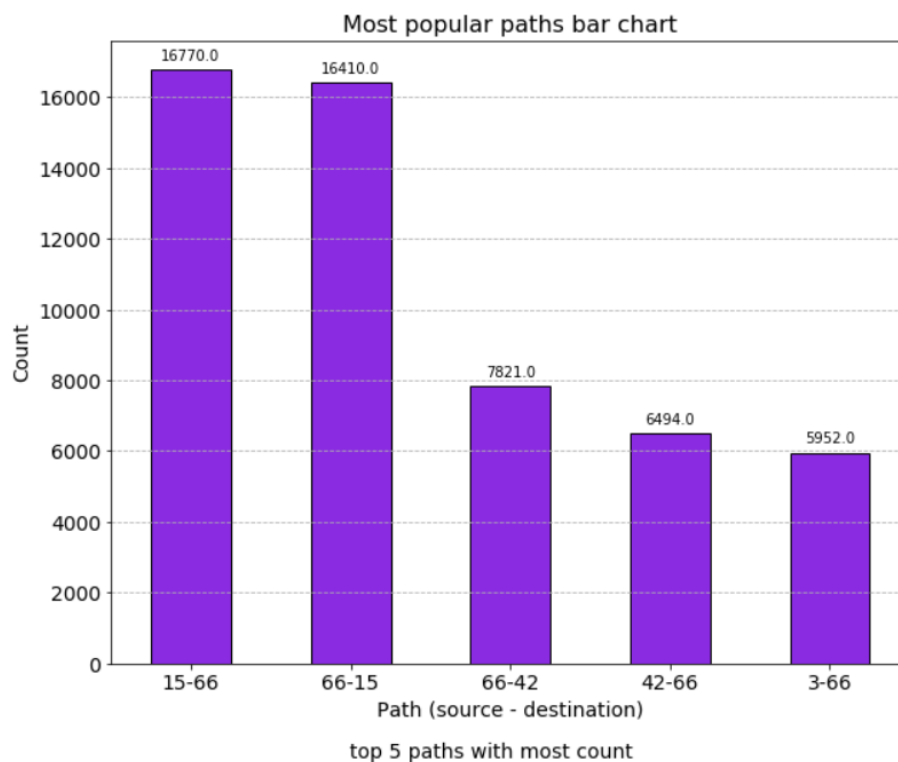
پرسش ۴- فاصله بین خرید و تاریخ بلیط معمولاً نمود خوبی از رفتار مشتری و همچنین اطمینان از سطح سرویس تامین کننده را به نمایش می گذارد. فاصله بین خرید و زمان حرکت را در ماه های مختلف بررسی کنید و آن را نمایش دهید. از این مقایسه چه نتیجه ای میگیرید؟

با توجه به نمودار ۴ میانگین بیشترین اختلاف بین تاریخ خرید بلیط و تاریخ پرواز مربوط به ماه February و کمترین مربوط به دو ماه April و October است. با توجه به میانگین اختلاف داده های گفته شده در هر ماه که بیشترین آن تقریباً ۹ روز است، نتیجه میگیریم که اکثر پروازها در ماهی که بلیط خریداری شده انجام شده است. این موضوع قابل انتظار بود چرا که با رسم نمودار ۵-۳ و مقایسه آن با نمودار ۴-۳ دیدیم که میزان خریده‌ها و پروازها در یک ماه متناسب است. می‌دانیم که بعد از ماه February تعطیلات ماه March است (نوروز). پس احتمال می‌رود افراد در ماه February بلیط‌های مسافرت تعطیلات March را خریداری میکنند تا بعداً به مشکل اتمام ظرفیت نخورند. این استدلال از نمودار ۵-۳ نیز قابل استناد است. در ماه January که میانگین کمتر است مشخص می‌شود که تعداد قابل توجهی سفر از نوع کوتاه مدت می‌باشد. در ماه‌های تابستانی نیز همچون May و June و July میانگین کم است. در شروع دوران تحصیلی در ماه August دوباره میانگین افزایش یافته است. پس احتمال می‌رود مردم بدلیل نگرانی از کمبود یا نبودن بلیط در ماه‌های نزدیک به ماه تعطیلات بلیط‌های خود را خریداری می‌کنند که این ممکن است به این علت باشد که مردم از سطح سرویس‌های تامین کننده در زمان تعطیلات اطمینان کمتری دارند اما در کل چون فاصله بین پروازها و خرید بلیت‌ها از مرتبه یک هفته است یعنی افراد به سرویس‌های تامین کننده اطمینان نسبی دارند.



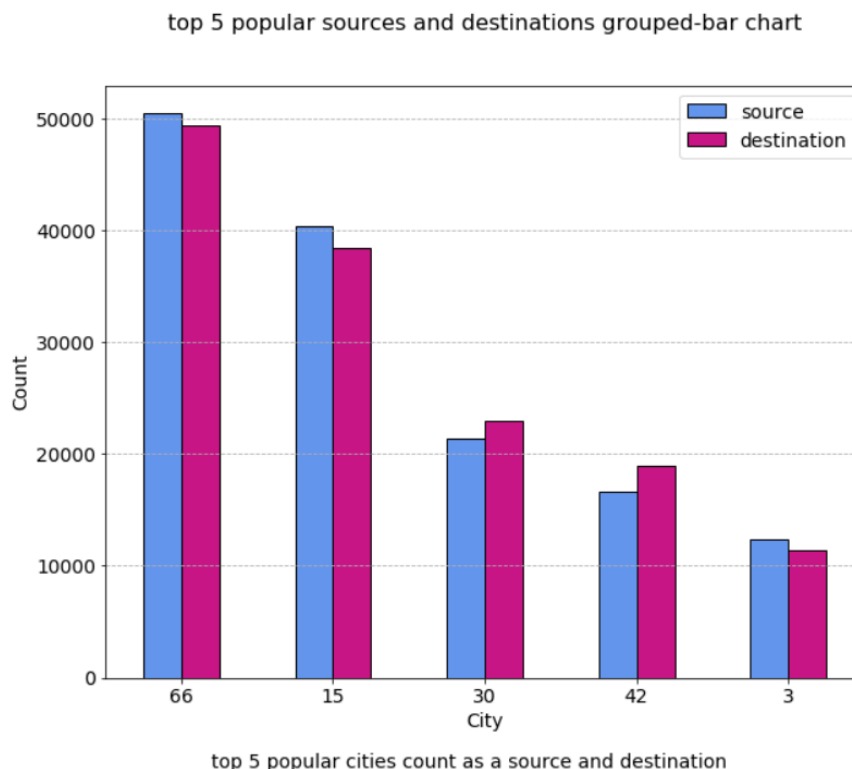
پرسش ۵ - پرطرفدارترین مسیرها کدام هستند؟ چه ارتباط بین پرطرفدارترین مسیرها و پرطرفدارترین مبداها و مقصدها وجود دارد؟

پرطرفدار بودن یک مسیر، مبدا و یا مقصد از میزان تکرار آن در بلیت های مسافران مشخص می شود. با توجه به نمودار کشیده شده برای ۵ مسیر پرتکرار، مسیر بین دو شهر ۱۵ و ۶۶ با مجموع ۳۳۱۸۰ بار استفاده پرطرفدارترین مسیر است.



نمودار ۱ - ۵

انتظار می‌رود با توجه به نمودار بالا شهرهای ۱۵ و ۶۶ در سه رده‌ی اول مبداها و مقصدها قرار داشته باشند. نمودار دوتایی پرتعدادترین مبداها و مقصدها در زیر آمده است:



نمودار ۲ - ۵

طبق انتظار شهرهای ۶۶ و ۱۵ هر دو در دو جایگاه اول قرار دارند اما برخلاف پیش‌بینی سومین مبدا/مقصد پرتعداد شهر ۳۰ است که اصلاً در ۵ مسیر محبوب ظاهر نشد. علت این موضوع برای کلان شهرها می‌تواند چنین باشد که افراد زیادی از تمام شهرهای دیگر برای انجام فعالیت‌های خود به این شهرها سفر می‌کنند و باز می‌گردند، در نتیجه به علت سفرهای زیاد به این شهر، طبیعی است که این شهر را در لیست مبداها و مقصدهای محبوب و پرتکرار ببینیم. اما از آنجا که مسافران از شهرهای مختلف به این شهر می‌آیند و سپس به شهر خود باز می‌گردند تعداد دفعات تکرار این شهر در دوتایی مبدا-مقصد بین شهرهای دیگر سرشکن می‌شود و می‌تواند لزوماً در بین ۵ مسیر پرتعداد قرار نگیرد. (یعنی به صورت فردی پرتکرار است اما لزوماً دوتایی یکتایی شامل آن شهر وجود ندارد که پرتکرار باشد). همچنین جمعیت زیاد یک شهر می‌تواند علت قرار گیری آن در مبداها و مقصدهای پرتعداد باشد. از نمودارهای بالا می‌توان استنتاج کرد شهرهای ۱۵ و ۶۶ نه تنها شهرهای صنعتی و بزرگی هستند بلکه جمعیت زیادی نیز دارند.

اما با اینکه یک مبدا/مقصد پرتکرار می‌تواند لزوماً در مسیرهای پرتکرار ظاهر نشود، مشاهده می‌کنیم که برعکس این رابطه برقرار است. یعنی اعضای یک زوج مرتب پرتکرار به صورت فردی نیز پرتکرارند. نمودار ۱-۵ را مشاهده کنید. نام شهرهای ۱۵، ۶۶، ۳ و ۴۲ در این دوتایی‌ها به چشم می‌خورد. حال نمودار ۲-۵ را مشاهده کنید. تمام این ۴ شهر در مبداها و مقصدهای پرتکرار وجود دارد. این امر منطقی است. هنگامی که زوج مرتب‌های مسیر سرشکن شده مربوط به یک شهر پرتکرار است، پس خود آن به صورت فردی باید بسیار پرتکرار باشد. از مقایسه اختلاف نمودارها نیز به نتایجی منطقی می‌رسیم.

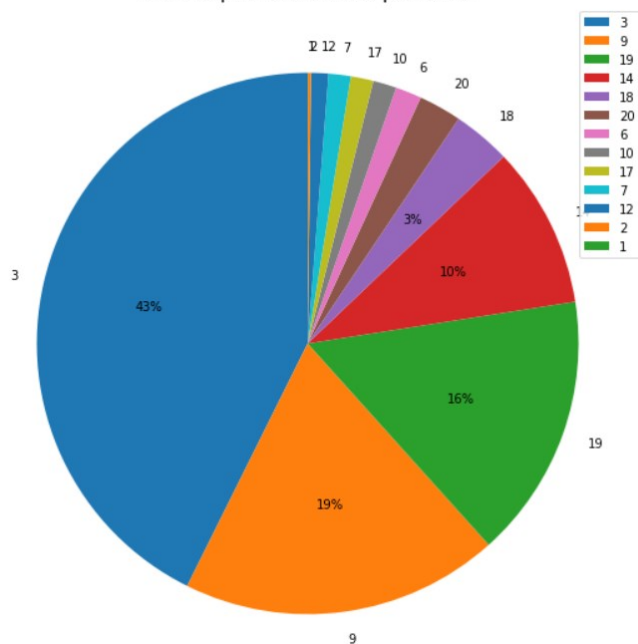
مسیر پرترفدار بین دو شهر ۱۵ و ۶۶ با اختلاف زیاد، اول است و دو شهر درگیر در این مسیر نیز با اختلاف زیاد نسبت به شهرهای سه مسیر پرترفدار دیگر اولند. همچنین شهر ۶۶ که در تمام مسیرهای پرترفدار تکرار شده است دارای بیشتر ارتفاع در مبدا و مقصد است.

پرسش ۶ - پرتعدادترین و پردرآمدترین ایرلاین‌ها کدام هستند؟ وابستگی این موضوع به ماه‌های سال و مبدا یا مقصد را بررسی کنید.

در این پروژه وابستگی پارامتر مورد نظر به مبدا را بررسی کرده‌ایم.

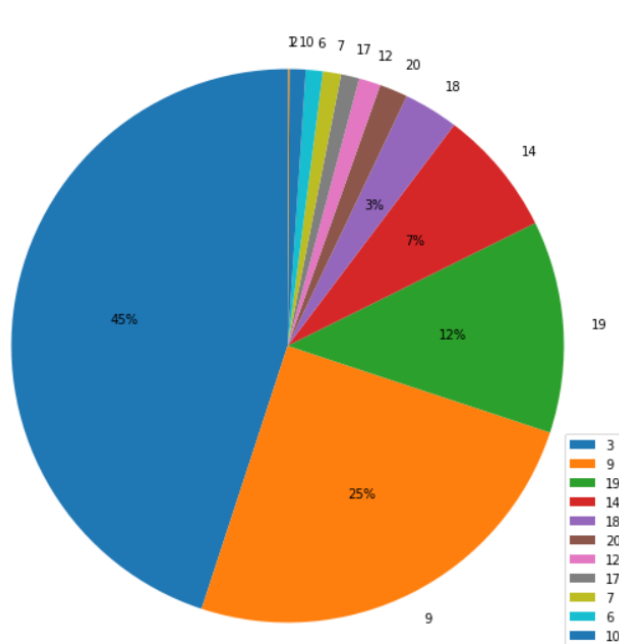
اگر قیمت‌های تمام ایرلاین‌ها تقریباً از یک مرتبه باشد انتظار می‌رود که پرتعدادترین ایرلاین (که ایرلاینی است که بیشتر از همه از آن سفارش گرفته شده است) پردرآمدترین ایرلاین نیز باشد. همچنین اگر یک ایرلاین با تعداد سفارش کمتر (محبوبیت کمتر) بخواهد پردرآمدترین باشد باید قیمت بلیت‌های آن مقدار قابل توجهی بیشتر باشد که همین امر نیز موجب کم شدن محبوبیت و تعداد سفارش‌های آن ایرلاین و نهایتاً کمتر شدن درآمد کل می‌شود. نتیجه نهایی این است که ایرلاین‌های پرتعداد، ایرلاین‌های پردرآمد نیز هستند.

All Companies incomes pie chart



نمودار ۲ - ۶: نسبت درآمد ایرلاین‌ها

All Companies popularity pie chart



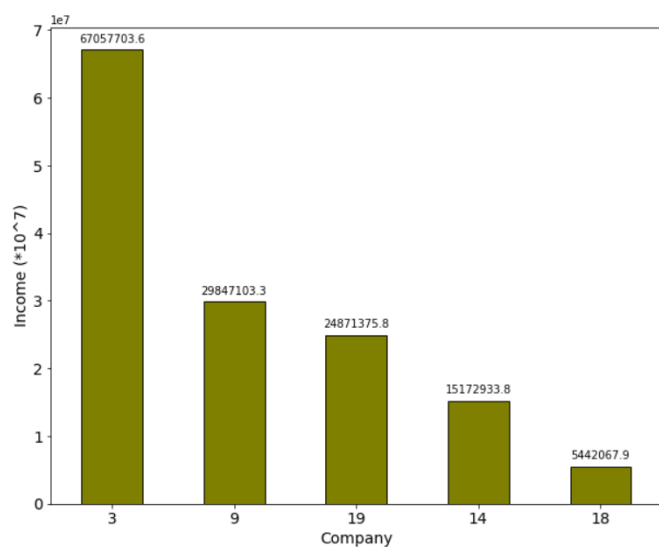
نمودار ۱ - ۶: نسبت محبوبیت ایرلاین‌ها

در نمودارهای بالا درآمد و محبوبیت ایرلاین‌ها مقایسه شده است. طبق انتظار میزان درآمد ایرلاین با میزان درخواست‌های آن (میزان محبوبیت) رابطه مستقیم دارد. در موارد آخر که قسمت‌های ایرلاین‌ها تقریباً برابر می‌شود اندکی جا به جایی در نمودار مشاهده می‌شود که می‌تواند بخاطر اندک اختلاف قیمت‌های آن‌ها باشد که به علت نزدیکی بسیار اعداد و اختلاف کم بین ایرلاین‌های ۶-۷-۱۲-۱۰-۲ چنین چیزی معمول است.

پروژهی تحلیل اطلاعات پرواز | ۲۴

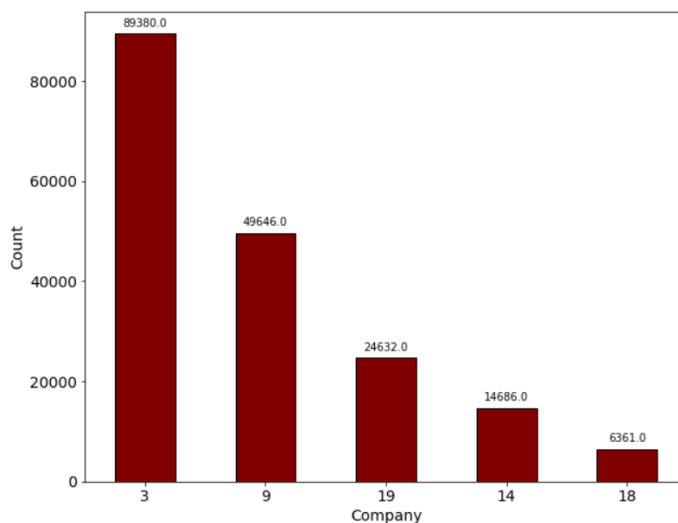
برای ۵ ایرلاین برتر در نمودارهای بالا، نمودار میله‌ای درآمد و تعداد درخواست در زیر آورده شده است:

top 5 high-income companies in total bar chart



top 5 companies with the most income
نمودار ۴ - ۶

top 5 popular companies in total bar chart

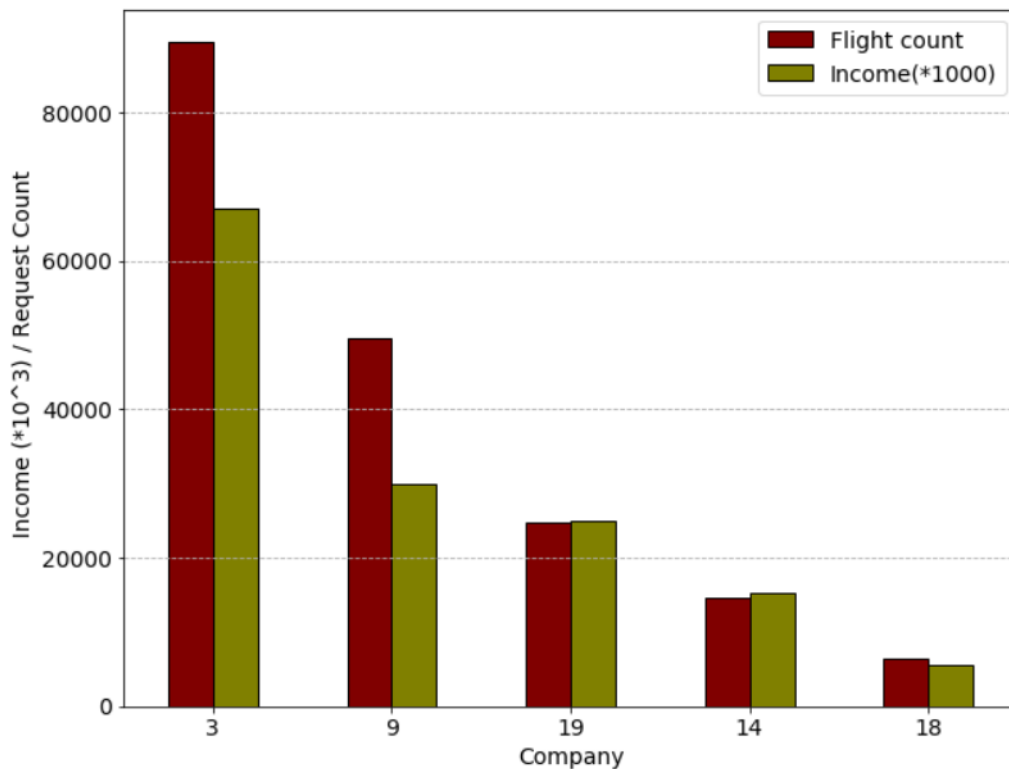


top 5 companies with most request count

نمودار ۳ - ۶

برای سهولت در مقایسه این دو نمودار را باهم نمایش می‌دهیم:

top 5 companies grouped-bar chart



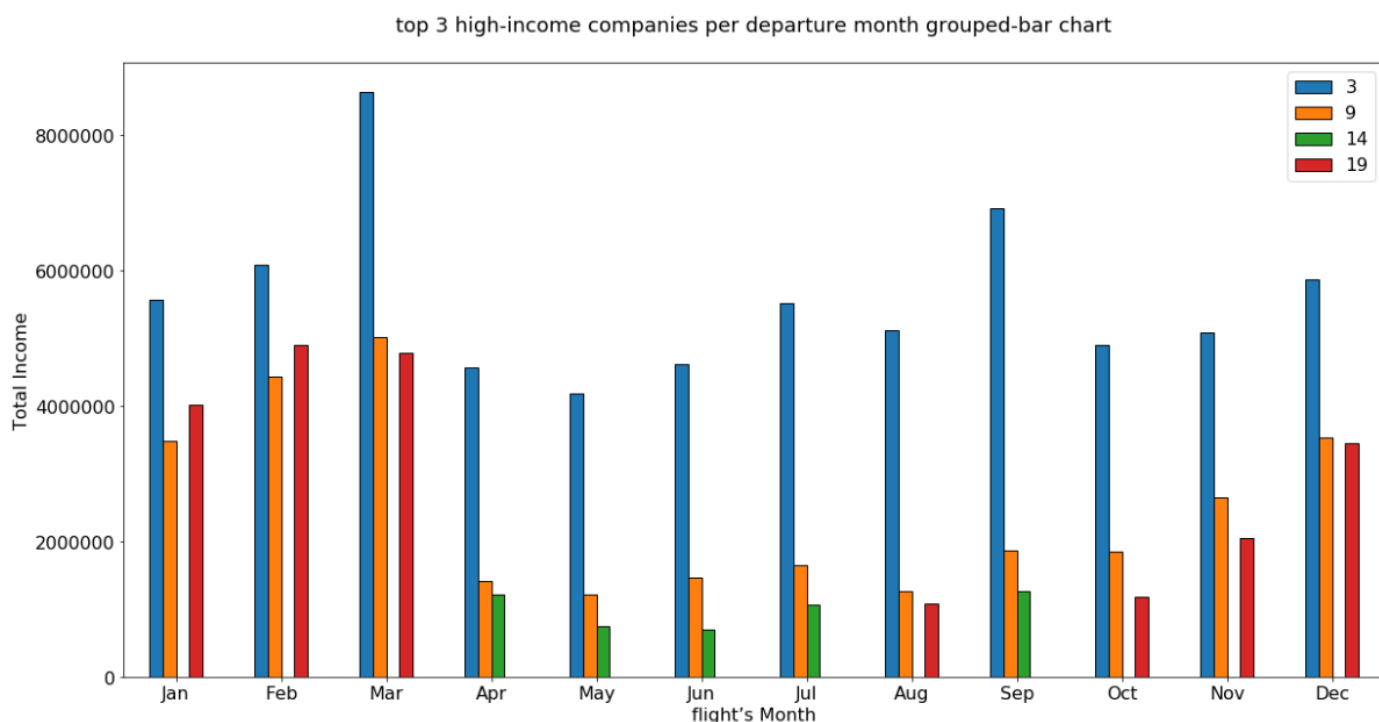
top 5 companies incomes and count

نمودار ۵ - ۶

همانطور که از نمودارهای دایروی برمی آمد هر دوی این نمودارها برای ایرلاین های ۳، ۹، ۱۹، ۱۴، ۱۸ به صورت نزولی هستند.

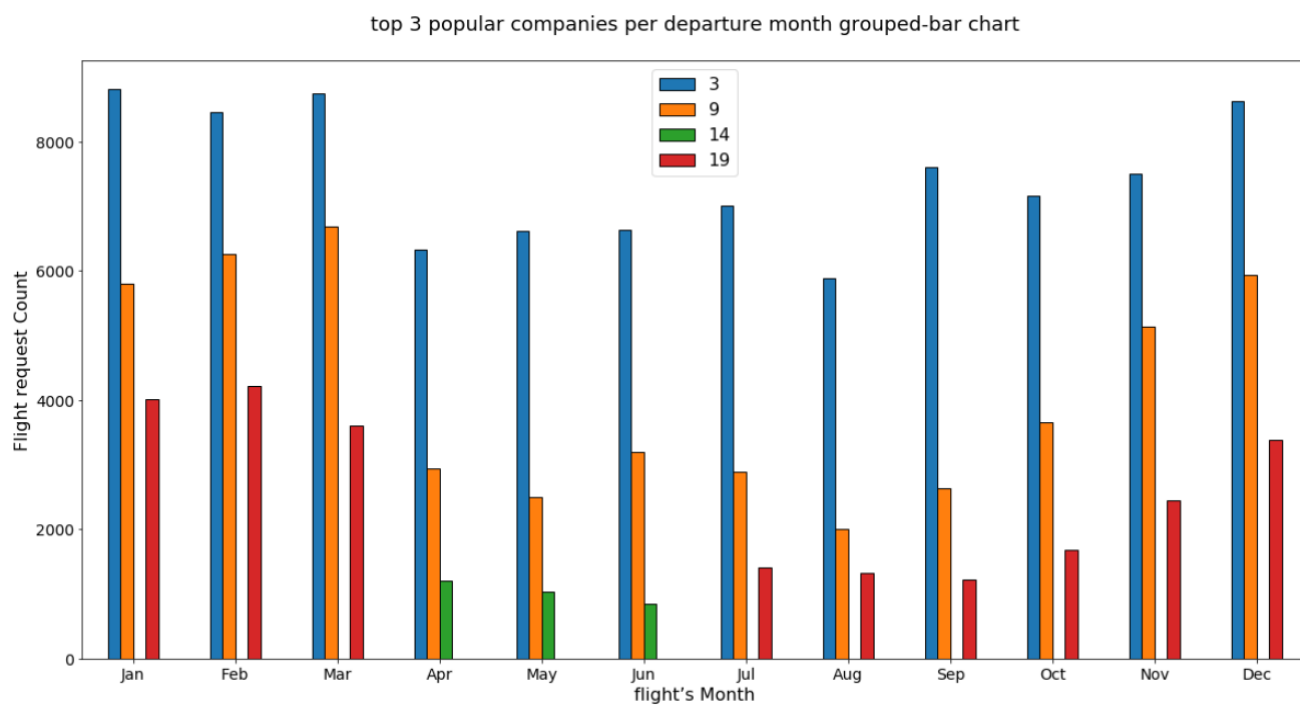
به نظر نمی رسد چنین موضوعی به مبدا یا ماه وابستگی داشته باشد مگر اینکه یک ایرلاین افزایش قیمت شدید یا افت کیفیت شدید (یا برعکس) داشته باشد یا مواردی از این دست. اما در حالت عادی و بدون تغییر ناگهانی انتظار می رود در هر ماه و در هر مبدا ایرلاین ۳ پردرآمدترین و پرتعدادترین ایرلاین باشد و همین ترتیب برای ایرلاین های ۹، ۱۹ و ۱۴ به صورت نزولی رعایت شود.

حال نمودار داده شده برای ماه ها را مشاهده می کنیم.



top 3 high-income companies in each month

نمودار ۶ - ۷

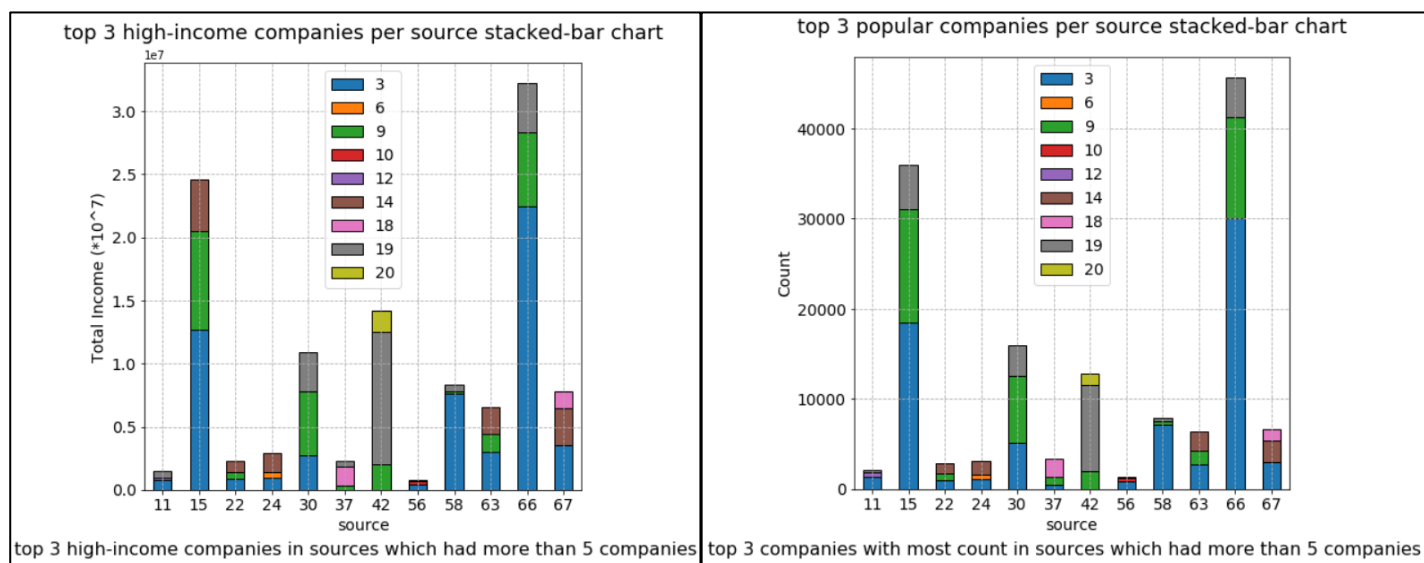


top 3 companies with most flight request count in each month

نمودار ۸ - ۶

دو نمودار بالا در هر ماه ۳ ایرلاین اول در درآمد و محبوبیت را نشان می‌دهند. طبق پیش‌بینی ایرلاین ۳ هم از لحاظ درآمد و هم از لحاظ محبوبیت در تمام ماه‌ها با اختلاف زیاد اول است سپس ایرلاین ۹ سپس ۱۹ و سپس ۱۴. در بعضی از ماه‌ها ایرلاین ۱۹ در سه ایرلاین برتر قرار نگرفته و ایرلاین ۱۴ جای آن‌را پر کرده است. در همان ماه‌ها درآمد ایرلاین ۱۴ از ۱۹ بیشتر بوده و ایرلاین ۱۹ باز هم در سه رده‌ی اول قرار نگرفته اما در ماه‌های سپتامبر و جولای درآمد ایرلاین ۱۴ از ایرلاین ۱۹ بیشتر بوده در حالی که درخواست‌های ایرلاین ۱۹ در همان ماه‌ها بیشتر از ۱۴ بوده. شاید بتوان گفت به این خاطر است که قیمت ایرلاین ۱۴ از ۱۹ بیشتر است اما با نزدیک شدن تعداد درخواست‌های این دو ایرلاین به یکدیگر (نمودارهای دایروی را ببینید. اندازه‌ی تکه‌ی این دو ایرلاین تقریباً هم اندازه‌اند) طبیعتاً ایرلاین ۱۴ درآمد بیشتری خواهد داشت. اما در اکثر ماه‌ها ترتیب پیش‌بینی شده رعایت شده و در کل می‌توان گفت رابطه مستقیم بین محبوبیت و درآمد برقرار است.

اکنون نمودارهای مربوط به مبدا هارا بررسی می‌کنیم.



نمودار ۱۰ - ۶

نمودار ۹ - ۶

نکته‌ی قابل توجه در این نمودارها وجود دارد و آن این است که تعدادی از مبداها در داده‌های استفاده شده از ۱ یا ۲ ایرلاین تنها استفاده کرده بودند (که آن یک یا دو ایرلاین نیز جزو همان ایرلاین های پرتعداد اول بودند). آوردن چنین مبدهایی برای مقایسه بین محبوبیت و درآمد ایرلاین‌ها معنی‌دار به نظر نمی‌رسد چرا که سوژه کافی برای مقایسه در آن‌ها وجود ندارد. از این رو در نمودار های ۹-۶ و ۱۰-۶ تنها مبدهایی آورده شده‌اند که بیش از ۵ ایرلاین در لیست سفارشات خود داشتند و ۳ ایرلاین برتر در هرکدام از آن‌ها به نمایش گذاشته شده‌است. همانطور که مشخص است در اکثر مبدها رنگ آبی که مربوط به ایرلاین ۳ است از همه بیشتر و با اختلاف بیشتر نسبت به بقیه رنگ‌ها به چشم می‌خورد. سپس رنگ سبز مربوط به ایرلاین ۹ و سپس طوسی مربوط به ایرلاین ۱۹. همچنین در تعدادی از مبدها (مانند ماه‌ها) رنگ قهوه‌ای مربوط به ایرلاین ۱۴ به جای طوسی (۱۹) دیده می‌شود که دلایل آن می‌تواند مرتبط با دلایل این اتفاق در نمودارهای ۶-۷ و ۶-۸ باشد و بهر حال در کل ایرلاین ۳ با اختلاف زیاد جایگاه اول را هم از نظر درآمد و هم از نظر محبوبیت در تمام مبدها دارد. شرایط مشابهی برای دوم بوم ایرلاین ۹ در بیشتر مبدها برقرار است. نتیجه نهایی این است که در کل محبوبیت و درآمد یک ایرلاین به ماه و مبدا (و انتظار می‌رود که مقصد) وابستگی ندارد. همچنین محبوبیت یک ایرلاین میزان پردرآمد بودن آن را مشخص می‌کند و برعکس. ایرلاین ۳ با اختلاف بسیار زیاد اولین ایرلاین هم از لحاظ درآمد و هم از لحاظ محبوبیت است و سپس ایرلاین‌های ۹، ۱۹ و ۱۴ به ترتیب دوم و سوم و چهارم (هم از لحاظ درآمد و هم از لحاظ محبوبیت) هستند.

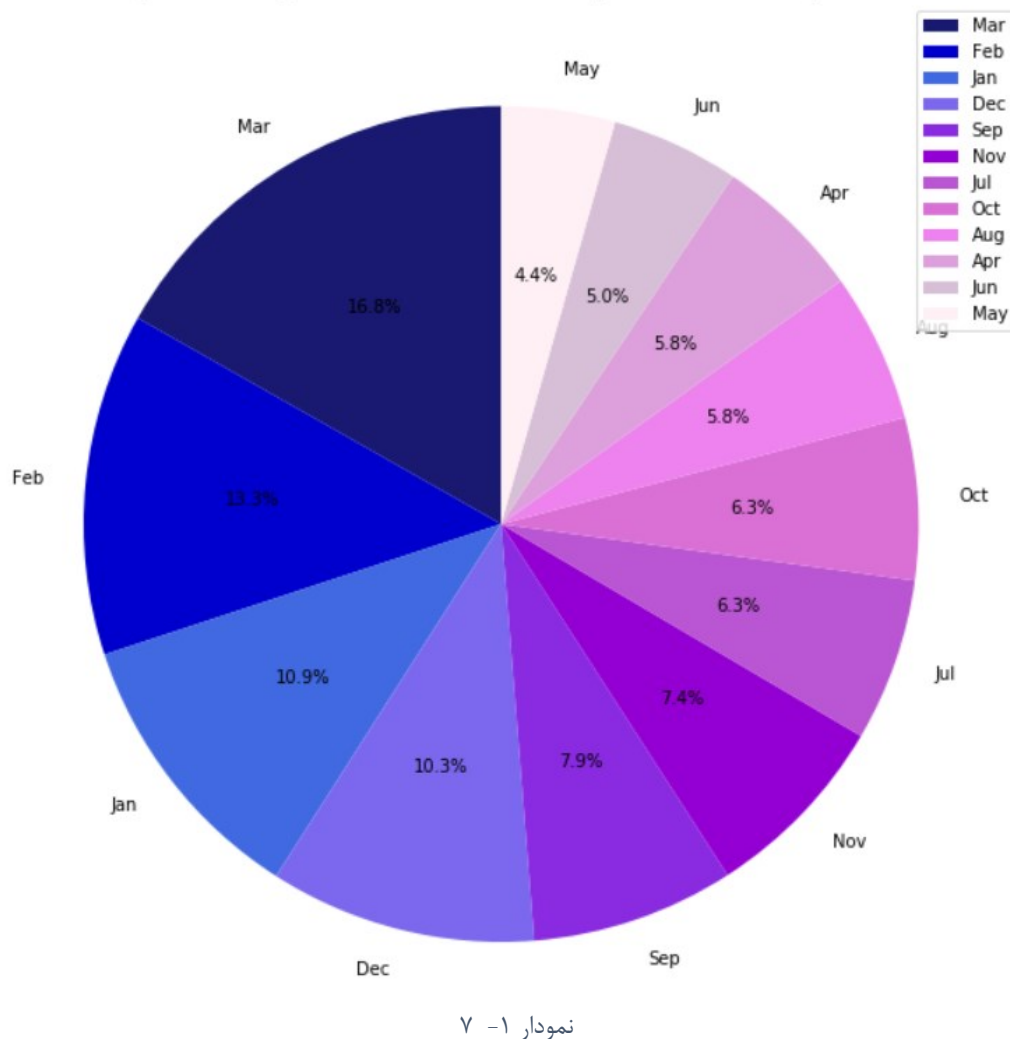
همچنین نگاهی سریع بیندازیم به ایرلاین‌های استفاده شده در شهرهایی که از کمتر از ۵ ایرلاین استفاده کرده بودند و در نمودار نیاوردیم. جدول روبه‌رو نام این ایرلاین‌ها و مجموع دفعات استفاده آن‌ها را نشان می‌دهد. می‌بینیم که در این جدول هم ایرلاین‌های ۳ و ۹ اول هستند و ترتیبی که بالاتر در محبوبیت ایرلاین‌ها دیدیم رعایت شده است.

| company | count | |
|---------|-------|---|
| 0 | 3 | 7 |
| 1 | 9 | 6 |
| 2 | 19 | 4 |
| 3 | 14 | 3 |
| 4 | 10 | 2 |
| 5 | 20 | 2 |
| 6 | 18 | 2 |
| 7 | 12 | 1 |
| 8 | 17 | 1 |

پرسش ۷ (پرسش جدید مطرح شده) - در کدام ماه‌ها در کل هزینه‌ی بیشتری برای بلیت هواپیما پرداخت شده است؟ تعداد پروازها در این ماه‌ها چقدر است؟ آیا نتیجه با نتیجه‌ی قابل انتظار تطابق دارد؟

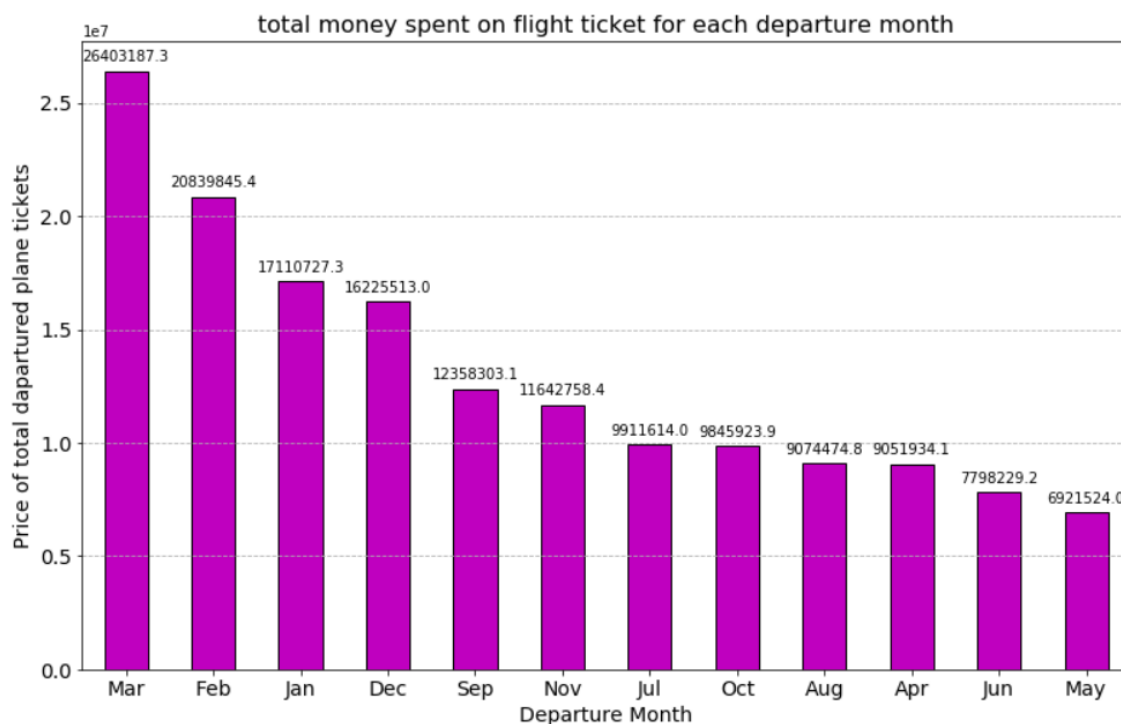
نتیجه‌ی قابل انتظار این هست که تعداد پروازها و به دنبال آن هزینه‌ی پرداخت شده برای پروازهای در ماه مارس بیشینه باشد چرا که یکی از مهم‌ترین تعطیلات - نوروز - در ماه مارس است و میزان سفر در تعطیلات بیشتر از بقیه زمان‌ها است.

total price of flight tickets for each departure month pie chart



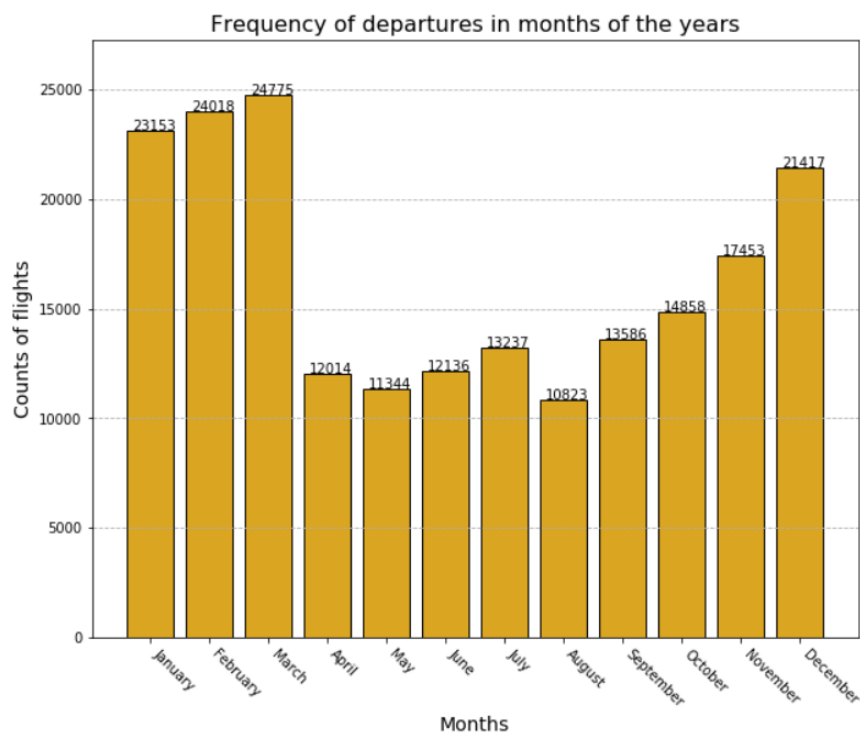
طبق انتظار هزینه‌ی صرف شده برای پروازهای ماه مارس که تعطیلات نوروز در آن قرار دارد با اختلاف، بیشتر از بقیه ماه‌هاست. بعد از آن ماه فوریه بیشترین هزینه پروازها را به خود اختصاص می‌دهد که می‌تواند به علت وجود تعطیلات ۲۲ بهمن در این ماه باشد. البته چون عمده تعطیلات مناسبتی به تاریخ قمری

بستگی دارند و در نتیجه در سال‌های مختلف تغییر می‌کنند تحلیل بیشتر این نمودار فقط برای بازه‌ی سال‌های جدول داده‌ها صحت خواهد داشت و در حالت کلی درست نخواهد بود.



نمودار ۲ - ۷

نمودار ۲-۷ نیز مقدار دقیق هزینه‌های کل پروازها در ماه‌های مختلف را نشان می‌دهد. کمترین مقدار هزینه‌های پرواز مربوط به ماه‌های می و ژوئن هستند که مصادف هستند با خرداد ماه که موقع امتحانات دانش‌آموزان و دانشجویان است و منطقاً انتظار می‌رود آمار سفر در این ماه‌ها به حداقل خود برسد.



نمودار ۵ - ۳

با توجه به نمودار هزینه‌های پرواز در ماه‌ها انتظار داریم که تعداد پروازها در ماه مارس بیشینه و در ماه می کمینه باشد. برای بررسی این موضوع نمودار ۵-۳ را دوباره در اینجا می‌آوریم:

طبق انتظار می‌بینیم تعداد پروازهای ماه مارس بیشینه است. اما اختلاف هزینه‌های بلیت در ماه مارس نسبت به ماه فوریه خیلی بیشتر از اختلاف تعداد بلیت‌های خریداری شده برای این دوماه است. (اختلاف ارتفاع بین ستون‌ها در دو نمودار قبل را مقایسه کنید). این در حالیست که در ماه‌های دیگر اختلاف ارتفاع بین ستون‌ها کم و بیش متناسب است. در ادامه علت احتمالی این موضوع را مطرح می‌کنیم.

در تعطیلات نوروز اغلب با افزایش قیمت هزینه‌های سفر از جمله هزینه‌های بلیت هواپیما مواجهیم. یعنی با وجود اختلاف کم بین تعداد پروازهای انجام شده در ماه‌های ژانویه، فوریه و مارس، اختلاف هزینه‌های این پروازها به دلیل گران شدن بلیت هواپیما با نزدیک شدن به نوروز بیشتر می‌شود. عکس این موضوع می‌تواند برای ماه می صادق باشد. ماه می از نظر تعداد پروازها، کمینه‌ی اول نیست. اما هزینه‌های آن کمینه است. شاید بتوان گفت به علت پایین آمدن میزان درخواست‌ها در فصل امتحانات برای سفر، ایرلاین‌ها قیمت خود را کاهش می‌دهند. این کاهش قیمت می‌تواند موجب تشویق به خرید و اندکی افزایش درخواست بشود. و نهایتاً ماه می با وجود اندکی پرواز بیشتر نسبت به آگوست، کمترین هزینه را داشته باشد. اما در کل تغییر بسیار شدیدی در ترتیب نزولی و صعودی بودن ماه‌ها در نمودارهای هزینه و تعداد پرواز دیده نمی‌شود و می‌توان گفت در کل هزینه‌های بلیت با تعداد پروازها در یک ماه مرتبط است (از آنجایی که در پرسش ۴ این مسئله را بیان کردیم که عموماً بلیت‌ها برای همان ماه خریداری می‌شوند) و به جز در زمان تعطیلات، جا به جایی‌های اندک، به علت نوسانات کم قیمت و نزدیکی تعداد بلیت‌ها بهم رخ می‌دهد.