

مقدمه

سلام. به پروژه پایانی بوت‌کمپ بهارانه تحلیل داده خوش آمدید. در این پروژه با همراهی تیم و با راهنمایی‌های منتور، داده‌های یک فروشگاه را بررسی می‌کنیم. بعد از مرتب‌سازی داده‌های مربوط با کمک آمار و یادگیری ماشین تحلیل‌های جذابی از این کسب و کار به دست خواهیم آورد و سپس این تحلیل‌ها را برای ارائه آماده خواهیم کرد.

کدهای مربوط به هر بخش را zip کنید و در آن بخش بارگذاری کنید. در صورتی که حجم فایل‌های شما بیش از ۲۰ مگ شد، فایل‌های جانبی را در درایوی با دسترسی عمومی قرار داده و لینک آن‌ها را در یک فایل .tex در فایل zip خود قرار دهید.

کدهای شما باید به‌خوبی داده‌هایی که در مراحل قبل به‌دست آورده‌اید را بازتولید کند، در غیر این‌صورت نمره‌ای از بخش‌های قبلی دریافت نخواهید کرد. همچنین در صورت وجود تقلب در کدها، طبق آیین‌نامه برخورد خواهد شد.

دقت کنید که آپلود فایل‌ها برای همه اعضای تیم الزامی‌ست.

ارائه پروژه روزهای سه شنبه ۳۰ خرداد و چهارشنبه ۳۱ خرداد خواهد بود. مانند پروژه قبلی حضور تمام افراد شرکت‌کننده در پروژه برای ارائه الزامی‌ست. عدم شرکت در ارائه به منزله‌ی عدم شرکت در پروژه حساب خواهد شد.

تمیزکردن و آماده‌سازی داده و پایگاه داده

داده اولیه این پروژه مربوط به یک فروشگاه است. شما در مرحله اول باید این دیتا را در پایگاه داده خود وارد کنید و در ادامه سوالات و خواسته های هر فاز را تکمیل کنید. برای نحوه وارد کردن دیتا به توضیحات زیر دقت کنید:

داده‌های اولیه را از اینجا دانلود کنید.

▼ توضیحات در مورد داده‌های اولیه

در فایل *Superstore.zip* فایلی به اسم *Superstore.sql* وجود دارد. با اجرای این فایل، همه جداول و سطرهایی که برای تست نهایی مورد استفاده قرار می‌گیرد در پایگاه داده شما ایجاد می‌شوند. پیش از اجرای این فایل، حتما پایگاه داده‌ای با اسم دلخواه ایجاد کنید و در ابتدای این فایل، کد `use YOUR_DB_NAME;` را وارد کنید.

در ادامه این فاز می بایست با توجه به دیتای اولیه موجود، به طراحی دیتاورهوس پرداخته شود. با توجه به آموزش های داده شده طراحی دیتاورهوس باید شامل جداول Fact و Dimension باشد و تا حد ممکن به ساختار star نزدیک باشد. وجود کلیدها و ارتباطات جداول از اهمیت زیادی برخوردار است.

مواردی که در این فاز مورد توجه است

- دریافت اطلاعات اولیه و پیش پردازش و تمیز کردن داده ها
- طراحی ساختار دیتاورهوس
- ایجاد کلیدها و ارتباطات جداول
- وارد کردن جداول در power bi و اطمینان از وجود ارتباطات درست جداول
- تمامی عملیات بالا باید توسط power bi انجام شود. در صورتی که از برنامه دیگری کمک گرفته شود ۸۰ درصد نمره‌ی سوال به تیم شما تعلق خواهد گرفت.

نکات و موارد امتیاز اضافی

- وجود اطلاعات جزئی تر در جداول Dimension عمومی مانند DimDate و DimGeography و
- افزودن اطلاعات بیشتر به جداول بر اساس خلاقیت افراد گروه به عنوان مثال اضافه کردن Longitude و Latitude به DimGeography و موارد مشابه.

- ایده های جدید و خلاقانه .

آمار

حتما تا الآن هر بار كه يك فروشگاه كه بر روى اجناس خود تخفيف زده است را ديده‌ايد، به اين فكر كرده‌ايد كه :«اين فروشگاهها با اين تخفيفاشون، ميان مردم رو گول ميزنن و يه عالمه جنس ميفروشن، تهشم بيشتر پول در ميارن». آيا به راستى اين حرف درست است؟

با توجه به داده‌هايى كه در اختيار شما قرار داده شده است، اين فرضيه را بررسى كنيد كه آيا تخفيف بر روى آيتم‌ها، باعث تغيير محسوسى روى فروش تعداد مى‌شود يا خير.

براى اين كار، شما بايد داده‌ها را به دو دسته‌ى تخفيف‌دار و بي‌تخفيف تقسيم كنيد. سپس، در هر دسته، توزيع تعداد آيتم‌هاى فروخته شده را بررسى كنيد، و سپس با روش‌هاى آمارى، بررسى كنيد كه تفاوت معنى‌دارى ميان دو دسته وجود دارد يا خير.

یادگیری ماشین

سوال اول

یکی از مهم‌ترین سوال‌هایی که در آنالیز و تحلیل شرکت‌های مختلف به کار می‌رود، تخمین سود به‌دست آمده است. یکی از کاربردهای تخمین سود، پیش‌بینی اتفاقات آینده و تصمیم‌گیری برای استراتژی‌های شرکت‌هاست.

در این بخش از شما خواسته شده است که با استفاده از داده‌هایی که در اختیارتان است، پیش‌پردازش آن‌ها و انتخاب ویژگی‌های مناسب، مدلی را آموزش دهید که سود یک محصول فروخته شده را محاسبه کند. به عبارت دیگر، شما باید مدلی طراحی کنید که با استفاده از تمام ویژگی‌های داده‌ها، ستون profit را پیش‌بینی کنید. برای این کار، می‌توانید از هر مدلی که به نظرتان مناسب است استفاده کنید. در این مسئله، به این موضوع دقت کنید که مدل شما قابل تعمیم (generalizable) باشد و دچار مشکل overfitting نشوید.

سوال دوم

یکی از راه‌های سود شرکت‌های فروش آنلاین، روش‌های حمل و نقل ویژه است. معمولاً شرکت‌ها سرویس‌های خاصی را به مشتریان می‌فروشند که در ازای پول بیشتر، اجناس را سریع‌تر ارسال می‌کنند. شرکتی که داده‌های آن در اختیار شما قرار داده شده است، می‌خواهد بداند که هر order، احتمالاً با چه ship mode انجام شده است. با داشتن این اطلاعات، شرکت می‌تواند با اندکی تخفیف درباره‌ی انواع مختلف حمل و نقل، مشتریان را تشویق کند تا به سراغ سرویس‌های گران‌تر بروند. برای مثال اگر یک order جدید قرار باشد به صورت استاندارد ارسال شود، شرکت می‌تواند تخفیف اندکی روی second class بگذارد تا مشتری تشویق شود و مدل حمل و نقل را یک مرحله بهتر کند. در این بخش شما باید پس از پیش‌پردازش داده‌ها و انتخاب ویژگی‌های مناسب، مدلی آموزش دهید که برای هر order، با داشتن تمامی ویژگی‌ها، حالت ارسال (Ship Mode) را پیش‌بینی کند.

داشبورد

در این بخش شما باید به طراحی داشبورد برای مدیر این کسب کار بپردازید و به اون پیشنهادهای لازم را بدهید. از هر روشی که بلدید، برای برای بهتر شدن و خواناتر شدن ارائه استفاده کنید. (مثلا اسم یا مقیاس محورها رو عوض کنید)

بخش اول

در بخش اول از داده‌هایی که از آمار و یادگیری ماشین به دست آمده استفاده کنید و گزارش‌های آنها را در اینجا بیاورید. توضیح مهم: دیتاها باید به داده های موجود در پایوری آی اضافه شوند.

بخش دوم

در بخش دوم به سوالاتی که مدیر از شما پرسیده است باید پاسخ بدهید که عبارت‌اند از :

- حجم هر بازار را (براساس میانگین و مجموع فروش) مشخص کنید و توضیح دهید سرمایه گذاری در کدام بازار منطقی‌تر است؟
- رابطه بین مبلغ سفارش و هزینه ارسال محصول را پیدا کنید. (می‌خواهیم بدانیم آیا کسانی که سفارش‌های گران‌تری دارند، برای ارسال محصول هم هزینه بیشتری پرداخت میکنند یا خیر؟)
- میانگین ارسال سفارش برای هر نوع ارسال در کشورهای مختلف چقدر است؟ (خودتون رو بذارید جای مدیر، برای هر منطقه -مثلا اروپا- هم باید به راحتی بتوانیم گزارش بگیریم.)
- در چه روزی از هفته (شنبه، یکشنبه، ...) فروش بیشتر بوده است؟
- فروشگاه از فروش چه محصولاتی بیشترین سود را کسب می‌کند؟ (هم بر اساس دسته‌بندی کلی و هم براساس هر جنس در هر دسته)
- معیاری تعریف کنید که اختلاف سود هر جنس با میانگین سود محاسبه کند و از این معیاری استفاده کنید تا متوجه شوید کدام منطقه سود بیشتری را برای فروشگاه به ارمغان آورده است؟

بخش سوم

در بخش سوم باید حداقل ۲ مورد به موارد بالا اضافه کنید تا بتوانید گزارش کاملی به مجموعه ارائه دهید.

راهنمایی

- اضافه کردن دیتاهای مورد نیاز از جاهای مختلف ممکن است به شما کمک کند.
- به این فکر کنید که مدیر از شما راهکارهایی میخواهد تا میزان فروش بالاتر رود.