

نحوه ی اجرای برنامه:

- 1- پایگاه داده: برای کار با پایگاه داده ابتدا یک پایگاه داده به نام `fpdb` میسازیم و سپس با استفاده از `pgadmin` بر روی پایگاه داده کلیک راست کرده و گزینه `rrestore` را انتخاب میکنیم و سپس فایل `db.tar` را از دایرکتوری پروژه انتخاب میکنیم.
- 2- برنامه C: با استفاده از دستور `pg_config --includedir` آدرس کتابخانه `libpq-fe` را میابیم و آن را `X` در نظر میگیریم. و سپس با دستور زیر کد را کامپایل میکنیم.
`gcc -o nameOfCompiledFile nameOfmainProgram.c -I/X -lpq -std=c99`
- 3- برنامه پاتون: برنامه پاتون را با دستور `python nameOfTheFile.py` اجرا میکنیم.

موارد بالا نحوه اجرای سه بخش اصلی پروژه است. ولی برای اجرای پروژه نیاز به انجام دستورات بالا نیست و تمامی دستورات بالا در فایل اسکریپت موجود در پروژه زده شده اند. برای اجرای پروژه کافی است که اسکریپت ها را اجرا کنید. و برای تنظیم زمان بندی ها از دستور `crontab -e` استفاده میکنیم و `cronjob` های زیر را درست میکنیم که خود به طور خودکار فایل های اسکریپت را اجرا میکنند.

```
* * * * * /Users/amin/Desktop/fifth-internship-project/fifth-internship-project/bash-files/download.sh
*/2 * * * * python /Users/amin/Desktop/fifth-internship-project/fifth-internship-project/bash-files/python.sh
```

پیاده سازی برنامه ها:

- 1 - برنامه C : این برنامه از دو بخش تشکیل شده است. یک بخش کار کردن با فایل و یک بخش کار با پایگاه داده. در ابتدا بوسیله یک تابع اطلاعات را از فایل ها میخواند و با استفاده از ۳ تابع اطلاعات را در درون سه تابع ذخیره میکنیم.
کتابخانه های مورد استفاده و موارد استفاده آنها:
`stdio`: برای کار با خواندن و نوشتن اطلاعات از فایل.
`stdlib`: برای تخصیص حافظه پویا.
`String`: برای مقایسه و `concatenaion` استرینگ ها.
`Libpq-fe`: برای کار با پایگاه داده.
`Dirent`: برای خواندن فایل های یک دایرکتوری.

- 2 - برنامه پایتون: در این برنامه داده ها را ۶ نوع تحلیل میکنیم. این کار را بوسیله پیاده سازی ۶ تابع انجام داده که در هر یک از تابع ها کویری های مربوط به تحلیل داده را انجام میدهیم. با اجرای این برنامه هر یک از این ۶ تابع به ترتیب فراخوانی میشوند.
کتابخانه های مورد استفاده و موارد استفاده آنها:
`math`: برای بدست آوردن ریشه دوم اعداد از این کتابخانه استفاده شده است.
`Psycopg`: برای کار با پایگاه داده.

توضیحات مربوط به تحلیل ها بر روی داده ها:

- 1 تحلیل اول: بیشترین محصولی که در هر استان مورد استقبال قرار گرفته است. این تحلیل را بر اساس تعداد کالای به فروش رفته از هر نوع انجام شده است. از نتیجه این تحلیل میتوانیم استفاده کنیم و کالاهای مورد علاقه در هر شهر را تامین کنیم(نگذاریم موجودی انبار به صفر برسد) و این که کالاهای مشابه را شناسایی کرده و در فروشگاه های آن شهر عرضه بکنیم.

به طور مثال خروجی کویری برای یکی از شهر ها
city = آزادشهر, product_id = 6125, has_sold = 39, total_money_earned = 1677000
می باشد که نشان میدهد شهر آزادشهر از محصول با شناسه ۶۱۲۵ به تعداد ۳۹ تا فروخته است و یعنی شهر آزادشهر از هیچ محصول دیگری به این تعداد نفروخته است.

2 تحلیل دوم : در این تحلیل بهترین فروشگاه را مشخص کرده ایم. معیار ما برای محاسبه بهترین فروشگاه مجموع پول بدست آمده از فروش کالا ها در هر فروشگاه می باشد. مقدار درآمد هر فروشگاه اطلاعات مطلوبی به ما نمیدهد اما با بررسی درآمد فروشگاه ها طی یک دوره میتوانیم فروشگاه هایی که در حال پیشرفت هستند و فروشگاه هایی که درامدشان در حال کاهش هست را بررسی کنیم و دلایل کاهش و افزایش درآمد را متوجه بشویم. دلایل کاهش درآمد به ما کمک میکند که از کاهش درآمد فروشگاه های دیگر پیشگیری کنیم و دلایل افزایش درآمد یک فروشگاه میتواند به ما کمک کند که درآمد فروشگاه های دیگر را نیز افزایش بدهیم.

به طور مثال خروجی حاصل از اجرای این تحلیل
market_id = 1511, income = 14875000
می باشد که نشان میدهد که فروشگاه با شناسه ۱۵۱۱ مجموع فروشش برابر ۱۴۸۷۵۰۰۰ بوده و از دیگر فروشگاه ها بیشتر است.

3 تحلیل سوم : در این تحلیل فروشگاه هایی که در آن ها کالاها با قیمت های گران تری به فروش رفته اند را مشخص کرده ایم. روش بدست آوردن به این صورت بوده است که به ازای هر کالا میانگین قیمت به فروش رفته ی آن کالا را در فروشگاه های مختلف محاسبه کرده ایم و به روشگاهی که بالاترین قیمت میانگین را برای آن کالا داشته است یک نمره منفی تعلق گرفته و در پایان که به ازای هر کالا این روند را تکرار کردیم آن فروشگاهایی که نمره منفی بیشتری گرفته است. فروشگاه گران فروشی محسوب میشود. از نتیجه آمار بالا برای برخورد کردن با فروشگاه های گران فروش میتوانیم استفاده بکنیم.

نتیجه ی یکبار اجرای این تحلیل بر روی داده ها
market with the id=106 is the most expensive market
market with the id=161 is the most expensive market
market with the id=938 is the most expensive market
market with the id=1102 is the most expensive market
market with the id=1673 is the most expensive market
market with the id=1201 is the most expensive market
market with the id=1368 is the most expensive market

می باشد که نشان می دهد ۷ تا از فروشگاه ها گران فروش ترین فروشگاه هستند یعنی این فروشگاه ها در تعداد برابری از کالا ها، کالاهای گران را به فروش گذاشته اند.

4 تحلیل چهارم: کالاهای که قیمت فروش متغیری داشته اند را مشخص کرده ایم. این تحلیل ممکن است برای بدست آوردن این که قیمت کدام کالاها بیشتر مستعد رشد است کاربرد داشته باشد. معیار مقایسه کالاها با یک دیگر انحراف معیار آن ها میباشد. هر کالا با قیمت های مختلفی به فروش رفته است که ما هر یک از آن قیمت ها را ثبت کرده و سپس انحراف از معیار آن ها را محاسبه میکنیم. هر کالا که انحراف از معیار بالاتری داشته باشد کالایی است که بیشتر مستعد رشد قیمت می باشد.

نتیجه ی یکبار اجرای تحلیل بالا بر روی داده ها
product with the id=7669 is the most variable product

میباشد که نشان میدهد قیمت های بفروش رفته برای کالای با شناسه ۷۶۶۹ انحراف معیار بیشتری دارند.

5 تحلیل پنجم : درآمد به طور متوسط فروشگاه های هر شهر را بدست آورده ایم. این تحلیل به ما کمک میکند که اگر خواستیم یک فروشگاه اضافه بکنیم شهر های که در آن ها فروشگاه ها فروش بیشتری داشته اند را بشناسیم و در آن شهر ها سرمایه گذاری بکنیم.
بطور مثال خروجی برای یک از شهر ها

average income of store in the city=خلیل آباد is 5278000

میباشد که نشان میدهد که متوسط درآمد فروشگاه های موجود در خلیل آباد برابر ۵۲۷۸۰۰۰ میباشد.

6 تحلیل ششم : کالاهایی که در هر یک از شهر ها بیشتر مورد نیاز است را پیدا کنیم. موجودی کالاها در هر یک از شهر ها بررسی میکنیم و همچنین میزان به فروش رفته ی هر یک از محصولات را نیز بررسی میکنیم. عدد به دست آمده از حاصل تقسیم عدد موجودی به عدد به فروش رفته ی هر کالا را به عنوان معیاری برای تاییین نیازمندیان به آن کالا در نظر گرفته ایم و کالایی که آن عدد برای آن نسبت به بقیه ی کالاها کوچکتر باشد به عنوان کالای بیشتر مورد نیاز در آن شهر معرفی میشود. این آمار برای توزیع کالا از انبار مرکزی به شهر ها کمک کننده است.

به عنوان مثال در نمونه زیر که برگرفته از بخشی از نتیجه ی اجرا این تحلیل میباشد

in the سردشت product with the id=9278 is most needed

در شهر سردشت کالای با شناسه ۹۲۷۸ از دیگر کالا ها در آن شهر بیشتر مورد نیاز میباشد(یعنی حال تقسیم میزان موجودی آن بر میزان به فروش رفته آن به نسبت دیگر کالا ها کوچکتر است).