پروژه درس داده‌کاوی – ترم پاییز 1401 – فاز 3 – دسته شماره 2

سید محمد طاها طباطبایی – ملیکا ذبیحی نیشابوری

**سوال اول : میزان نرخ ورود دارایی به تفکیک سال، کالا**

در این سوال خواسته شد، تا میزان کالای ثبت شده در سیستم دارایی برحسب هر سال، محاسبه شود. برای انجام اینکار به این صورت عمل کردیم که، برحسب ویژگی CREATED در تیبل PRODUCTS که تاریخ ثبت یک کالا را در سیستم نشان می‌دهد، سال ثبت کالا در سیستم را استخراج و به تعداد کالاهای ثبت شده در آن سال اضافه می‌کنیم. با یکبار پردازش تیبل، سال ثبت هر کالا مشخص می‌شود.

در ابتدای کد، با بررسی تیبل(سورت تاریخ ها)، متوجه شدیم که قدیمی ترین کالا مربوط به سال 2013 و جدید ترین کالا در سال 2022 ثبت شده است. این دو مقدار را به عنوان کران لیست تاریخ ها در نظر گرفتیم.

تابع goodsPerYearCounter ، برای استخراج سال ثبت کالا و آپدیت تعداد کالاهای ثبتی هر سال استفاده می‌شود. در این تابع به ازای هر رکورد، فرمت تاریخ بررسی می‌شود(ممکن است در ستون CREATED دیتا غیر معتبر داشته باشیم)، سپس اگر این رکورد یک تاریخ معتبر داشت، بخش سال این تاریخ در نظر گرفته می‌شود و در لیست تاریخ ها، به فراوانی اندیس متناظر با این سال، یکی اضافه می‌کنیم.

در ادامه کد، یک لیست با طول کران های مشخص شده (2013 – 2022) می‌سازیم، و تابع goodsPerYearCounter را فراخوانی می‌کنیم.

در نهایت مقادیر لیست نرخ کالای ورودی را چاپ می‌کنیم.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| سال | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| تعداد کالا | 21033 | 6079 | 4599 | 3974 | 3591 | 1555 | 1035 | 1175 | 1031 | 2168 |

**سوال دوم: بررسی ورود کالا و مقایسه تعداد و ارزش آن با معاونت و هلدینگ**

در این بخش، از ما خواسته شد تا به ازای هر هلدینگ، بررسی کنیم تا این هلدینگ چه نوع کالاهایی دارد و از هر نوع کالا، چه تعداد و به چه ارزشی موجود است. در اینجا، ما پردازش ها را روی تیبل PRODUCTSINSTANCE انجام دادیم. در این تیبل، نوع کالا را با آیدی آنها(ستون M\_PRODUCT\_ID) متناظر در نظر گرفته ایم، یعنی کالا هایی با آیدی یکسان، از یک نوع هستند. در فایل Dictionary نیز این ویژگی با اسم شناسه کالا عنوان شده. برای محاسبه مجموع ارزش کالا ها نیز، از ستون BOOKVALUE استفاده کردیم که طبق توضیحات، قیمت روز کالا را مشخص می‌کند.

برای تقسیم بندی کالا ها، از دو ساختار داده دیکشنری استفاده می‌کنیم. ایده این است که دیکشنری اول، کالا ها را بر اساس هلدینگ سیو کند، و دیکشنری دوم بر اساس اینکه چه کالایی در چه هلدینگی قرار دارد، تعداد و قیمت آن را نگهداری کند.

ساختار دیکشنری اول که با نام holdingInfo تعریف می‌شود، به شکل زیر است:

{ key:"ACCT\_AC\_HOLDING\_ID" , value:"ItemKey" }

کلید این دیکشنری، هر آیدی از ستون ACCT\_AC\_HOLDING\_ID است که آیدی هر هلدینگ است.

مقدار value این دیکشنری نیز، یک رشته است که از ترکیب مقدار آیدی هلدینگ و آیدی کالا ساخته می‌شود. با این تعریف، کالا-هلدینگ را می‌سازیم. این رشته را به اسم ItemKey می‌شناسیم که به عنوان کلید دیکشنری دوم استفاده می‌شود. اگر این کلید را به دیکشنری دوم بدهیم، یک تاپل باز می‌گرداند که تعدا و قیمت بر اساس نوع کالا در هلدینگ است.

ساختار دیکشنری دوم:

{ key:"ItemKey" , value:"(items count , items cumulative value)" }

دیکشنری دوم در کد با اسم itemInfo تعریف می‌شود.

برای پاسخ به سوال دوم، ابتدا تیبل PRODUCTINSTANCE را لود می‌کنیم. در ابتدا بخش cleaning دیتا را داریم. در بخش ابتدا باید در ستون ACCT\_AC\_HOLDING\_ID مقادیر غیر از آیدی را حذف کنیم. چون برخی از رکورد های این ستون شامل داده های رشته ای بود، با تابع stringRemover مقادیر رشته ای را که امکان تبدیل شدن به داده عددی دارند، به عدد تبدیل می‌کنیم، و بقیه موارد تبدیل به null می‌شوند.

سپس دو ستون M\_PRODUCT\_ID", "BOOKVALUE" را با خروجی مرحله قبل کانکت می‌کنیم تا یک دیتافریم شامل 3 ستون مورد نیاز داشته باشیم. در انتها تمام رکورد های null را حذف می‌کنیم تا دیتا های مناسب پردازش باقی بمانند.

در بخش پردازش دیتا، ابتدا در یک حلقه، بر اساس هلدینگ ها پردازش را شروع می‌کنیم. آیدی هلدینگ را در متغییر holdingId نگهداری می‌کنیم. به ازای هلدینگ فعلی، مقدار آیدی کالا و ارزش آن را نیز ذخیره می‌کنیم. سپس رشته ItemKey را می‌سازیم. در ادامه بر اساس چند شرط، دیکشنری های خود را آپدیت می‌کنیم. در شرط اول، بررسی می‌کنیم، آیا این هلدینگ تا کنون در دیکشنری ما ثبت شده است یا نه. در صورت پاسخ منفی، هلدینگ به انتهای دیکشنری اول اضافه می‌شود و مقادیر تعداد و قیمت نیز به انتهای دیکشنری دوم اضافه می‌شوند. اگر آن هلدینگ در دیکشنری موجود باشد، شرط دوم بررسی می‌کند که آیا کالای در حال بررسی تا کنون در این هلدینگ بوده یا نه. اگر نباشد، به انتهای دیکشنری اول اضافه می‌شود و تاپل آن نیز ساخته می‌شود. اگر باشد، فقط تاپل مربوط به آن در دیکشنری دوم آپدیت می‌شود.

در انتهای کد و در تابع printFirstNItem، به طور نمونه، مقادیر تعداد و ارزش چند کالا در هلدینگ های مختلف را چاپ می‌کنیم. خروجی نمونه به این شکل است:

ItemID\_HoldingID: 104011689\_1

count: 459 value: 208992.43 valueSum: 95927527

در اینجا نتیجه را اینطور تفسیر می‌کنیم که آیتم از نوع ItemID در هلدینگ HoldingID دارای فراوانی count، ارزش واحد value و مجموع ارزش valueSum است.

**سوال سوم: لیست اموالی که سند حسابداری نخورده اند و بررسی آنها که مثلا پیش‌نویس یا نهایی شده اند**

در این سوال خواسته شده تا اموال را بر اساس اینکه در چه وضعیتی هستند، دسته‌بندی کنیم. برای اینکار، دو ستون "C\_DOCSTATUS\_ID","INOUT\_ID" از تیبل INOUT را انتخاب می‌کنیم تا هر کالا را بر اساس وضعیت سند آن دسته‌بندی کنیم.

برای دسته بندی، از اعمال شرط روی دیتافریم استفاده کرده ایم. به طول کلی 3 حالت برای سند ممکن است، حالت 3000006 که حالت پیش‌نویس است. حالت 3000025 که حالت ثبت نهایی است، و وضعیت هایی غیر از این دوحالت که همگی را در یک دسته سوم دسته‌بندی می‌کنیم.

نتیجه به شکل دو ستون چاپ می‌شود که آیدی و وضعیت هر سند را نشان می‌دهد.