# "بسمه تعالى"

# گزارش پروژه Cube گزارش

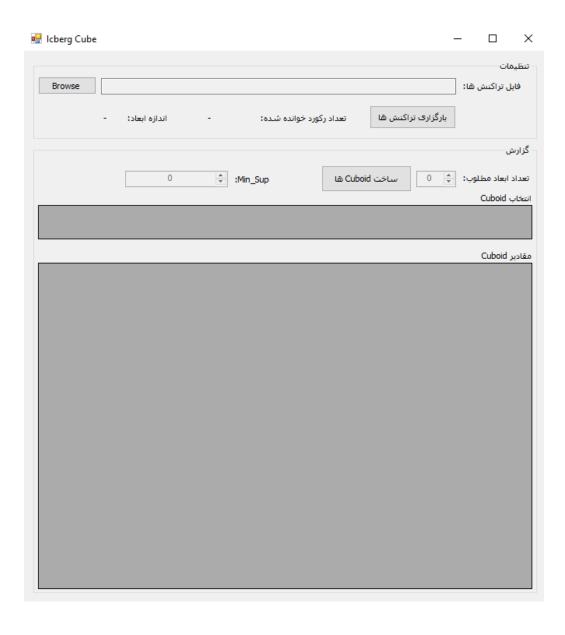
درس: داده کاوی

استاد: جناب آقای دکتر احمدی

تهیه کننده: فرخنده زینالی آق قلعه

شماره دانشجویی: ۹۶۱۱۲۷۴

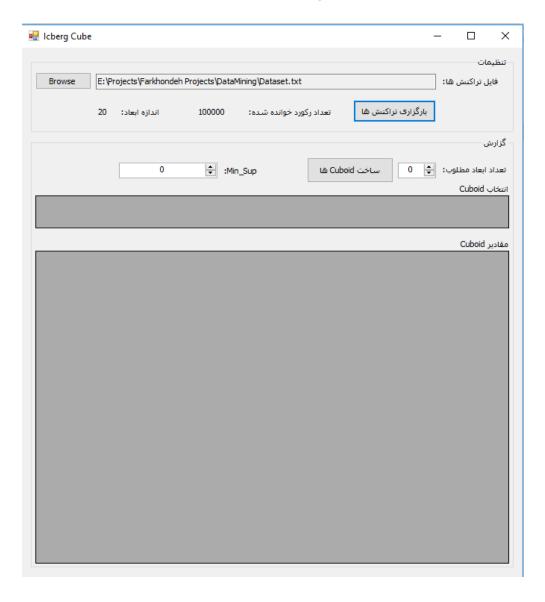
در این پروژه قصد داریم از الگوریتم Iceberg Cube برای یک dataset که شامل ۱۰۰۰۰۰ تراکنش از ۲۰ آیتم مختلف که از ۰ تا بعد از ۱۹ شماره گذاری شده اند، تعداد مشاهده زیرمجموعه های آیتم های مذکور را به عنوان Aggregate نمایش دهیم. لذا بعد از پیاده سازی الگوریتم مذکور می توانیم Cuboid های nبعدی و نتیجه Aggregate را برای هر یک از آنها مشاهده کنیم. جهت پیاده سازی این الگوریتم یک برنامه ویندوزی با استفاده از نرم افزار 2010 Microsoft Visual Studio و به زبان #C و روی Net Framework 4 نوشته است که فرم نهایی این برنامه را در زیر مشاهده می کنید.



# مراحل ساخت پروژه

# ۱- خواندن و بارگزاری dataset

همانطور که در تصویر مشاهده می کنید با استفاده از دکمه Browse ابتدا مسیر فایل dataset را در برنامه انتخاب می کنیم و با زدن دکمه بارگزاری تراکنش ها عملیات بارگزاری انجام خواهد شد. پس از اینکه بارگزاری با موفقیت انجام شد برنامه تعداد رکورد های خوانده شده و اندازه ابعاد یعنی تعداد آیتم های متمایز خوانده شده را در لیبل های مربوطه نمایش می دهد که در اینجا به ترتیب برابر ۱۰۰۰۰۰ تراکنش و ۲۰ آیتم می باشد.

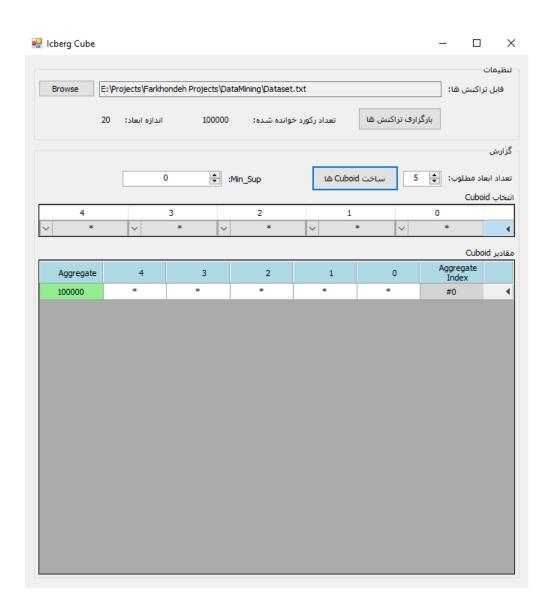


#### ۲- ساختن Cuboid ها

پس از خواندن تراکنش ها و تعیین شدن تعداد کل آنها و اندازه ابعاد موجود در آنها، range گزینه های دو فیلد تعداد ابعاد مطلوب از ۰ تا اندازه ابعاد و Min\_Sup از ۰ تا تعداد کل تراکنش ها تعیین می شود که تنها در این Rin\_Sup اعداد معتبر خواهند بود. سپس جهت ساخته شدن Cuboid ها می بایست تعداد ابعاد مطلوب خود را انتخاب کنیم و دکمه ساخت ها را بزنیم.

# **توجه**: در این برنامه انتخاب اندازه مطلوب و ساخت Cuboidها به صورت <u>داینامیک</u> تعیین می شود که از مزیت های برنامه محسوب می شود.

از آنجایی که صورت مسئله تعداد ابعاد مطلوب را ۵ معرفی کرده است ما نیز فیلد تعداد ابعاد مطلوب را ۵ انتخاب می کنیم و دکمه ساخت Cuboidها را می زنیم. در نتیجه پس از اتمام عملیات تعداد ۵ آیتم که از ۰ تا ۴ اندیس گذاری شده اند جهت انتخاب Cuboid ها خواهیم داشت. در زیر نمایش فرم برنامه را پس از ساخت Cuboidها مشاهده می کنید.

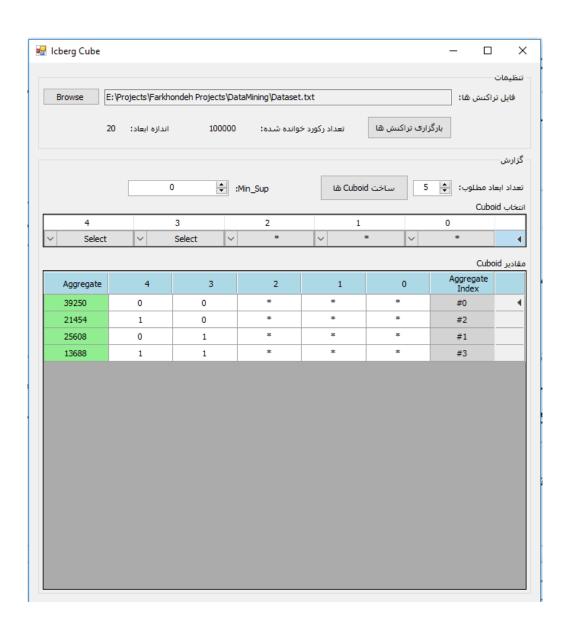


#### ٣- مشاهده نتىجه

پس از تعیین تعداد ابعاد مطلوب و ساخت Cuboid در قسمت آنتخاب Cuboid می توان به ازای هر یک از ابعاد مطلوب حضور یا عدم حضور را تعیین کرد در ابتدا همانطور که مشاهده کردید تمام ابعاد ۰ تا ۴ با علامت \* به معنای عدم حضور تنظیم شده اند. در صورتی که بخواهیم وضعیت حضور هر یک از ابعاد را تغییر دهیم می توانیم combobox مربوطه به آن بعد را به Select تغییر دهیم.

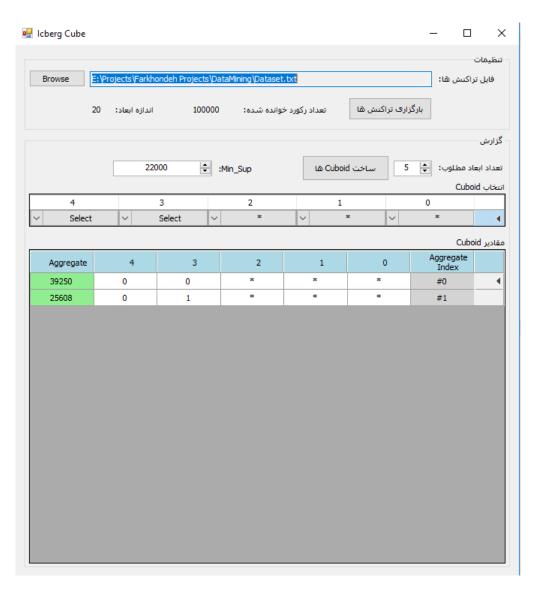
همانطور که در تصویر قبل مشاهده می کنید در ابتدا تمام ابعاد وضعیت \* یا عدم حضور را دارند و در نتیجه بعد ۰ مد نظر بوده و به اصلاح Apgregate انتخاب شده است که در قسمت مقادیر Cuboid تنها یک Aggregate موجود می باشد و تعداد آن نیز برابر تعداد کل تراکنش ها یعنی ۱۰۰۰۰۰ است که با رنگ سبز متمایز شده است.

با تغییر وضعیت حضور ابعاد در قسمت انتخاب Cuboid نتیجه Aggregate ها در قسمت مقادیر Cuboid نیز به صورت اتوماتیک تغییر می کند. نتیجه انتخاب ابعاد چهارم و پنجم یعنی ابعاد شماره ۳ و ۴ را در تصویر زیر مشاهده می کنید.



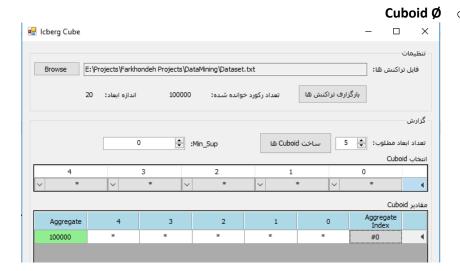
همانطور که در تصویر بالا مشاهده می کنید از آنجایی که دو بعد انتخاب شده اند لذا چهار Aggregate برای ترکیب حضور یا عدم حضور هر یک از بعد ها به دست می آید. هر یک Aggregate Index را با Index# در ستون Aggregate Index که با رنگ خاکستری متمایز شده اند مشاهده می کنید.

جهت فیلتر کردن Aggregate ها نیز می توان از فیلد Min\_Sup استفاده کرد که مقدار پیش فرض آن برابر ۰ می باشد که در این صورت تمامی Aggregate ها نمایش داده خواند شد ام با افزایش مقدار Min\_Sup آن دسته از Aggregate هایی که مقدار آنها از Min\_Sup کمتر باشد حذف شده و فقط مقادیر بزرگتر نمایش داده خواهند شد. در زیر نمایش Aggreagte های انتخاب دو بعد شماره ۳ و ۴ با Min\_Sup برابر ۲۲۰۰۰ را مشاهده می کنید که از میان ۴ Min\_Sup تنها ۲ تای آنها که بزرگتر از مقدار Min\_Sup داشته اند نمایش داده شده است.



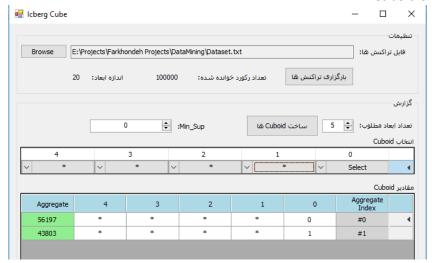
در زیر مقادیر Aggregate برای تمامی وضعیت های حضور یا عدم حضور ابعاد شماره ۰ تا ۴ آورده شده است:

• بعد •

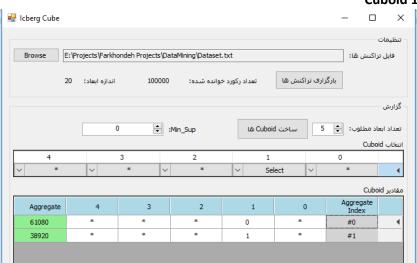


# • بعد ١

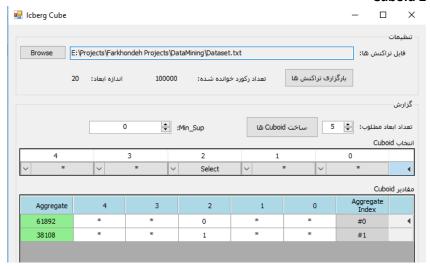
#### Cuboid 0



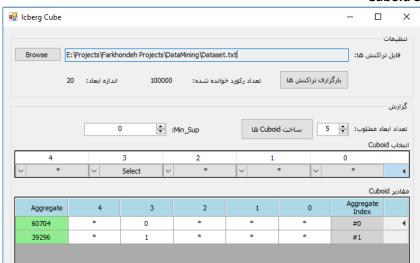
# Cuboid 1 o



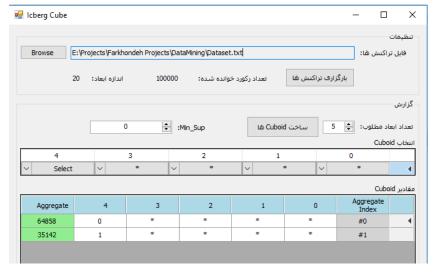
# Cuboid 2 o



#### Cuboid 3 o

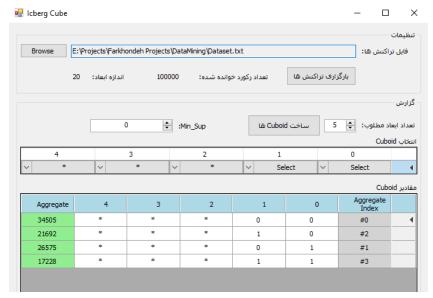


# Cuboid 4 o

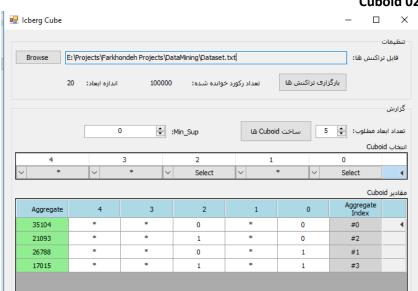


# • بعد ۲

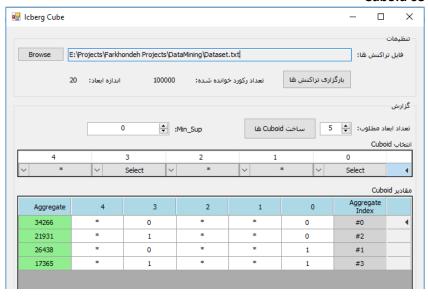
#### Cuboid 01 o



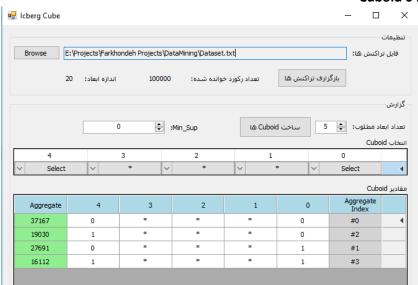
#### Cuboid 02 o



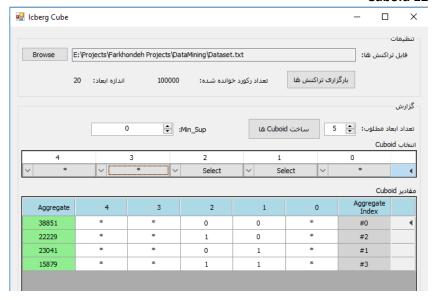
#### Cuboid 03 o



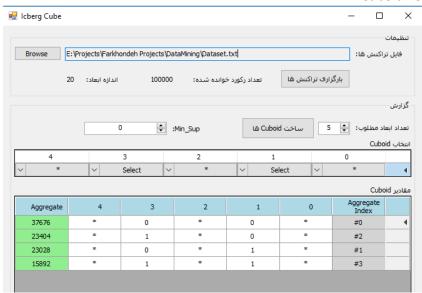
# Cuboid 04 o



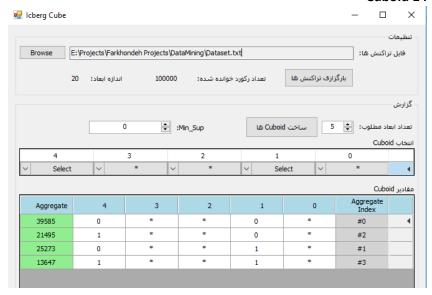
#### Cuboid 12 o



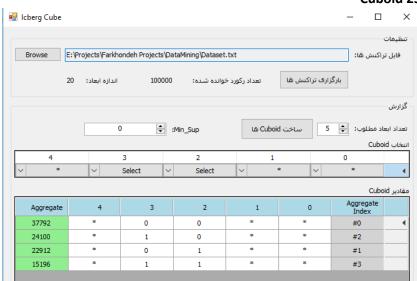
# Cuboid 13 o



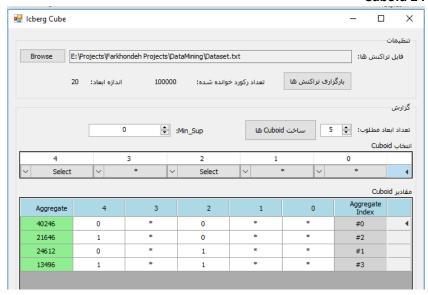
# Cuboid 14 o



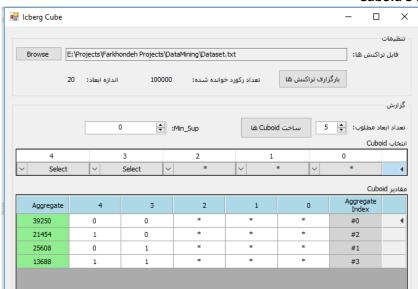
#### Cuboid 23 o



#### Cuboid 24 o

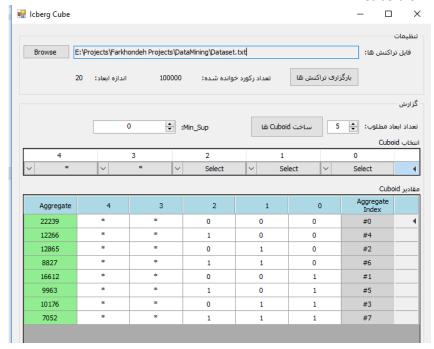


#### Cuboid 34 o

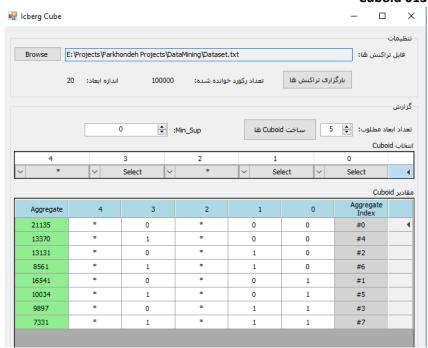


# • بعد ٣

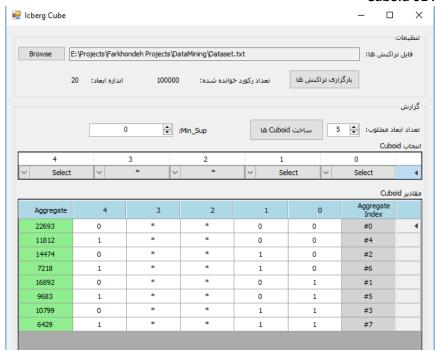
# Cuboid 012 o



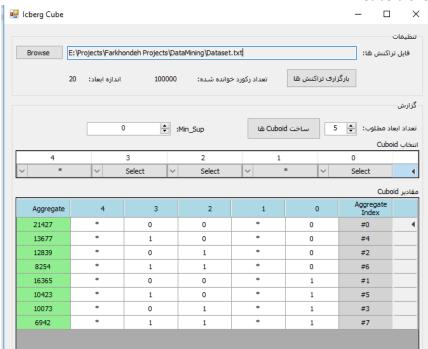
#### Cuboid 013 o



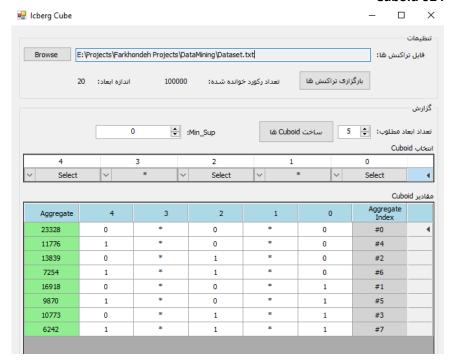
#### Cuboid 014 o



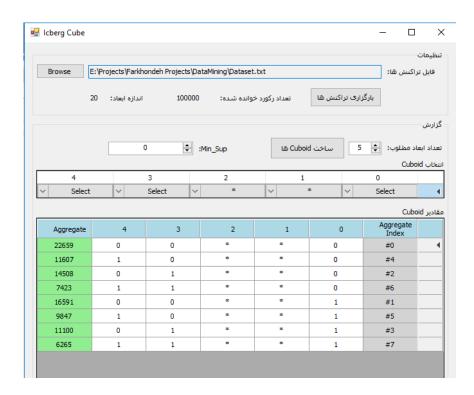
#### Cuboid 023 o



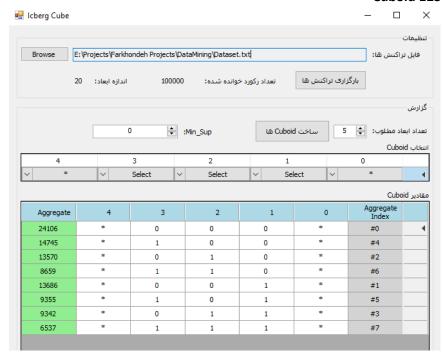
#### Cuboid 024 o



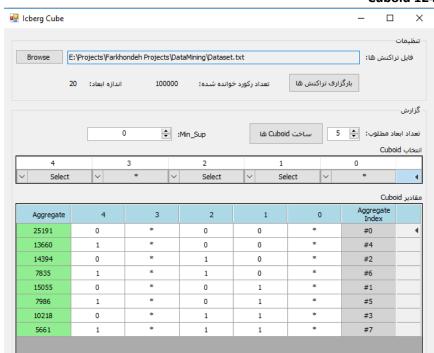
#### Cuboid 034 o



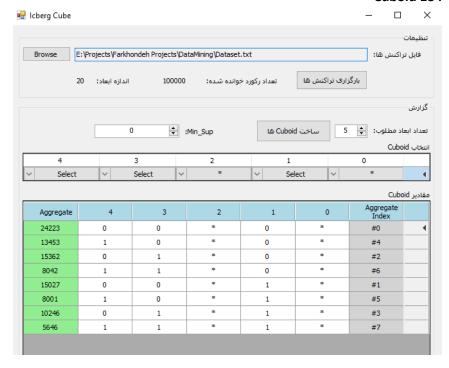
#### Cuboid 123 o



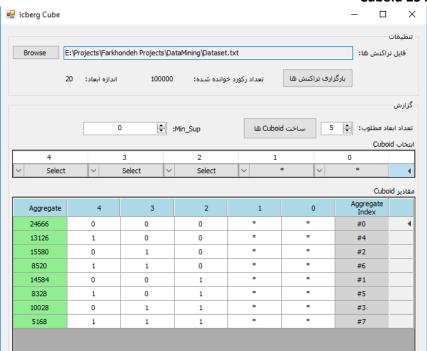
#### Cuboid 124



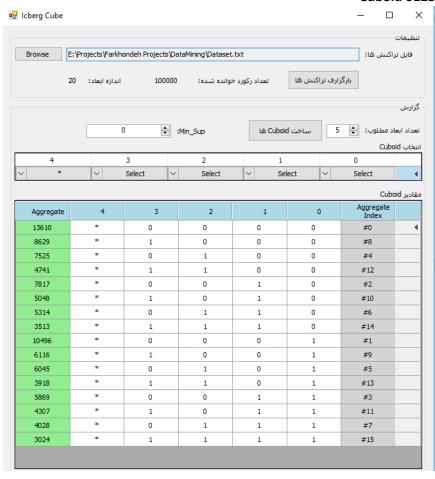
# Cuboid 134 o



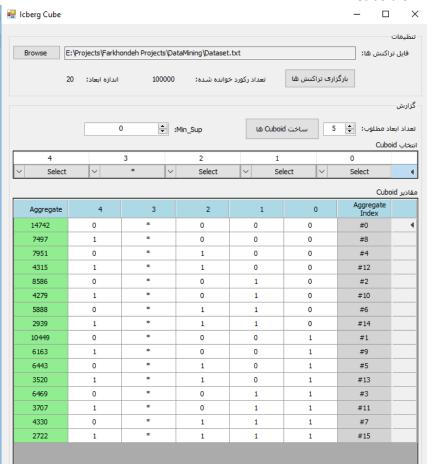
# Cuboid 234 o



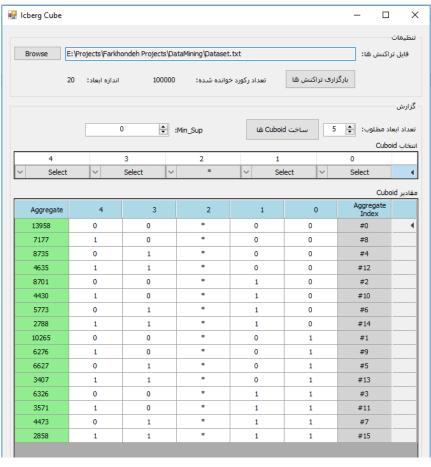
#### **Cuboid 0123** o



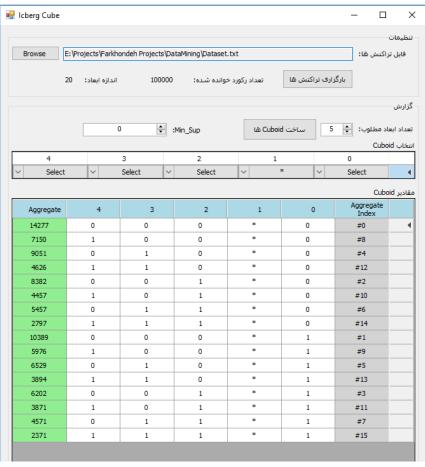
# **Cuboid 0124** o



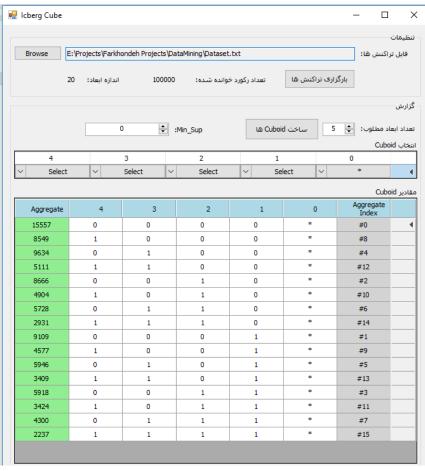
#### **Cuboid 0134** o



#### **Cuboid 0234** o



#### **Cuboid 1234** o



# • بعد ۵

#### **Cuboid 01234** o

