



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مدیریت علم و فناوری

گزارش کار هفته دوم

تفسیر مصورسازی داده ها

نگارش
رضا اکبری مقدم

استاد
دکتر مهدی قطعی

آبان ماه ۹۹

فهرست

۴مقدمه
۴Data Sets
۴مجموعه داده بیمارستان عمومی هورتون
۴مجموعه داده اماکن گردشگری عراق
۵مجموعه داده سهام شرکت البرز
۶نرم افزار weka
۹نرم افزار orange
۱۲نرم افزار Power BI
۱۴نرم افزار Rapid Miner
۱۴منابع

فهرست اشکال

شکل ۱.....	۶
شکل ۲.....	۶
شکل ۳.....	۷
شکل ۴.....	۷
شکل ۵.....	۸
شکل ۶.....	۹
شکل ۷.....	۱۰
شکل ۸.....	۱۱
شکل ۹.....	۱۲
شکل ۱۰.....	۱۳
شکل ۱۱.....	۱۴

مقدمه

ضمن عرض سلام و خسته نباشید خدمت استاد محترم، در این تمرین بنده به تحلیل و تفسیر انواع data sets در ۴ نرم افزار گوناگون پرداختم و نتایج این تحقیق به شرح زیر میباشد:

Data Sets

مجموعه داده بیمارستان عمومی هورتون

این مجموعه داده در سال ۲۰۱۹ بر روی بیمارستان هورتون انگلستان بدست آمده است، که اطلاعاتی را در رابطه بیماران بخش اورژانس این بیمارستان میدهد.

این مجموعه داده دارای ۶ ویژگی عددی شامل:

- Adm : تعداد بیماران بخش اورژانس.
- Cardio : تعداد بیماران اورژانسی که ایست قلبی کرده‌اند.
- Resp : تعداد بیماران اورژانسی که ایست تنفسی کرده‌اند.
- Hypo : تعداد بیماران اورژانسی که کاهش قند خون داشته‌اند.
- Month : ماه (۱-۱۲)
- Year : سال (۱۹۹۹-۲۰۱۱)

همچنین این مجموعه داده شامل ۱۳۹ نمونه میباشد

مجموعه داده اماکن گردشگری عراق

این مجموعه داده در سال ۲۰۲۰ بوجود آمده که آماری در ارتباط با توریست های کشور عراق در شهرها و اماکن مختلف آن نشان میدهد.

این مجموعه داده دارای ۶ ویژگی عددی شامل:

- ID(int) : توریست
- City(text) : شهر
- place names(text) : نام مکان
- latitude(real) : عرض جغرافیایی
- longitude(real) : طول جغرافیایی
- C (boolean) : فرهنگی

- E (boolean) : محیطی
- H (boolean) : پزشکی
- N (boolean) : طبیعت
- R (boolean) : مذهبی
- B (boolean) : کاری
- L (boolean) : تفریحی

این مجموعه داده شامل ۲۳۲ نمونه میباشد.

مجموعه داده سهام شرکت البرز

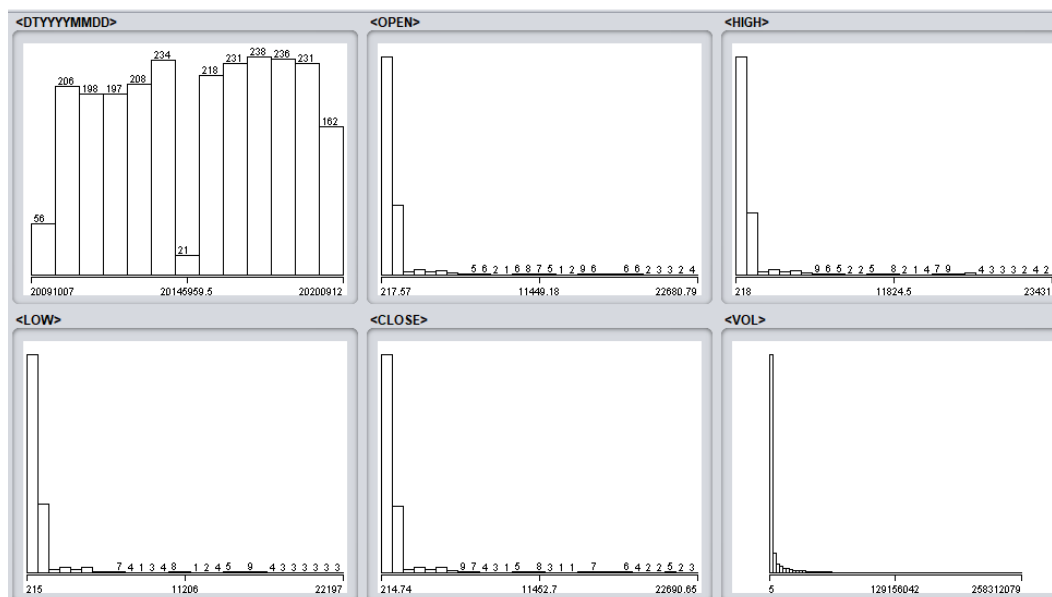
این مجموعه داده اطلاعات قیمتی و معاملاتی سهام شرکت البرز را نشان میدهد.

این مجموعه داده شامل ۶ ویژگی عددی میباشد :

- DTYYYYMMDD : تاریخ
- Open : قیمت باز شدن
- High : بالاترین قیمت
- Low : کمترین قیمت
- Close : قیمت پایانی
- Vol : حجم معاملات

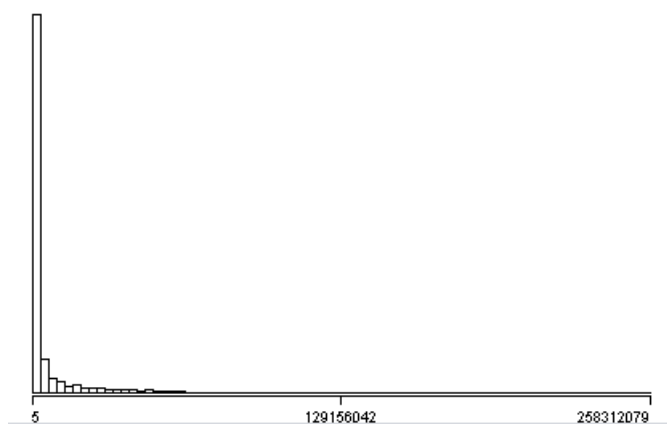
این مجموعه داده شامل ۲۳۴۷ نمونه میباشد.

نرم افزار weka



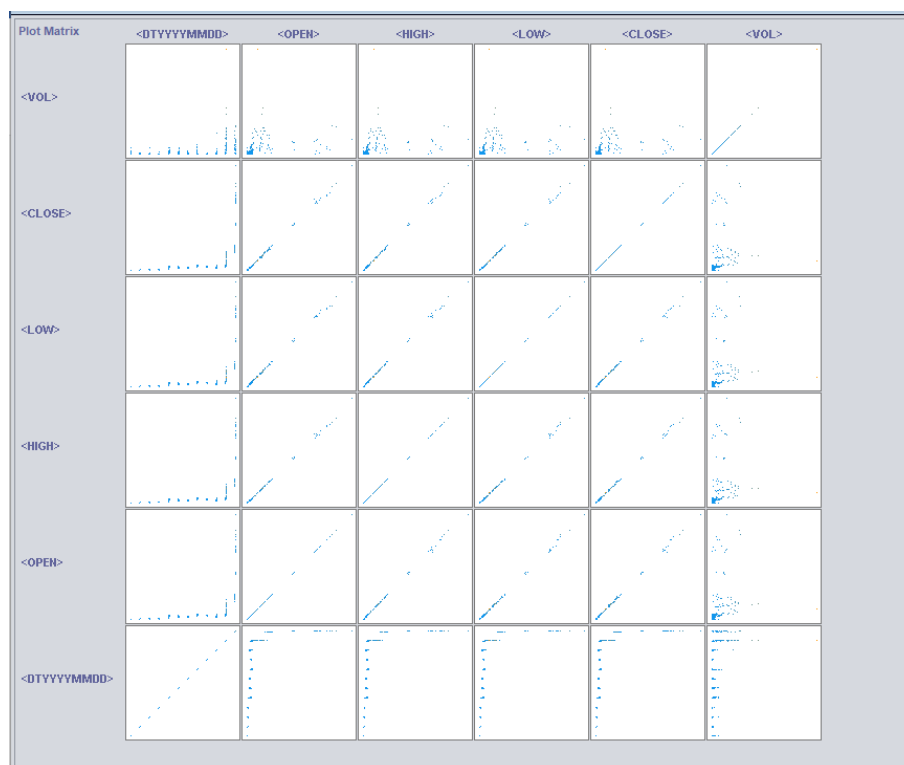
شکل ۱

در شکل ۱ نمودار میله ای هر ویژگی جداگانه نمایش داده شده است که نشان میدهد در هر ویژگی در چه مقادیری تجمع نمونه ها بیشتر میباشد.



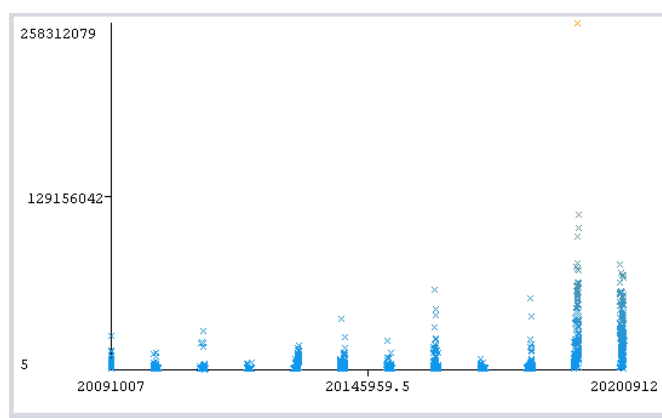
شکل ۲

برای مثال در نمودار شکل ۲ حجم معاملات نشان میدهد بیشترین حجم معاملات در ناحیه ۵ بوده است.



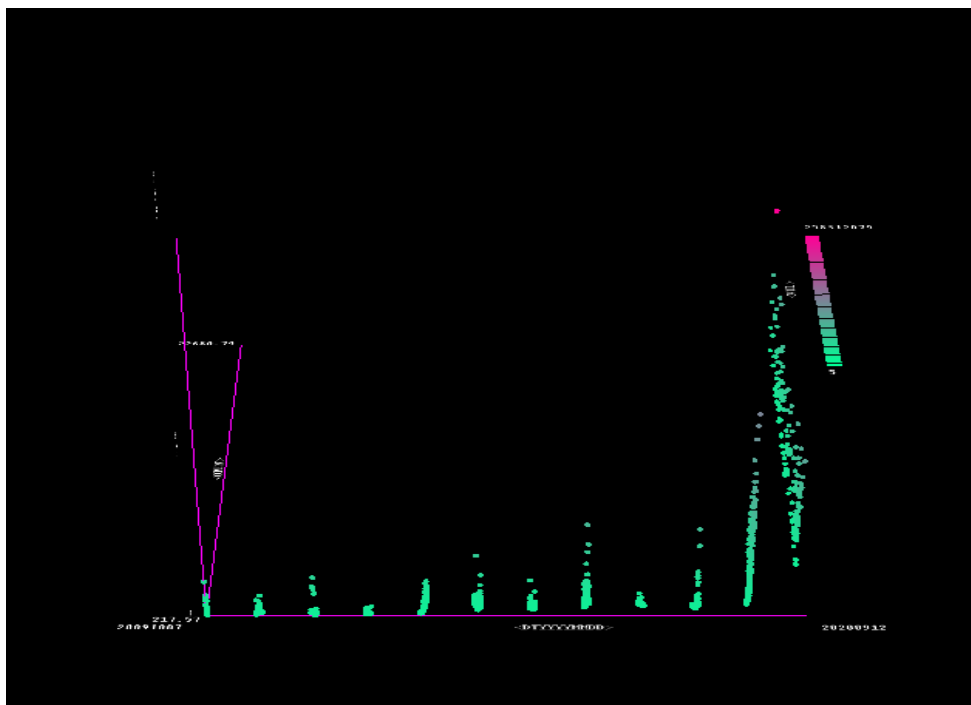
شکل ۳

همچنین با استفاده از این نرم افزار می‌توان Scatterplot Matrices را نمایش داد. (شکل ۳)



شکل ۴

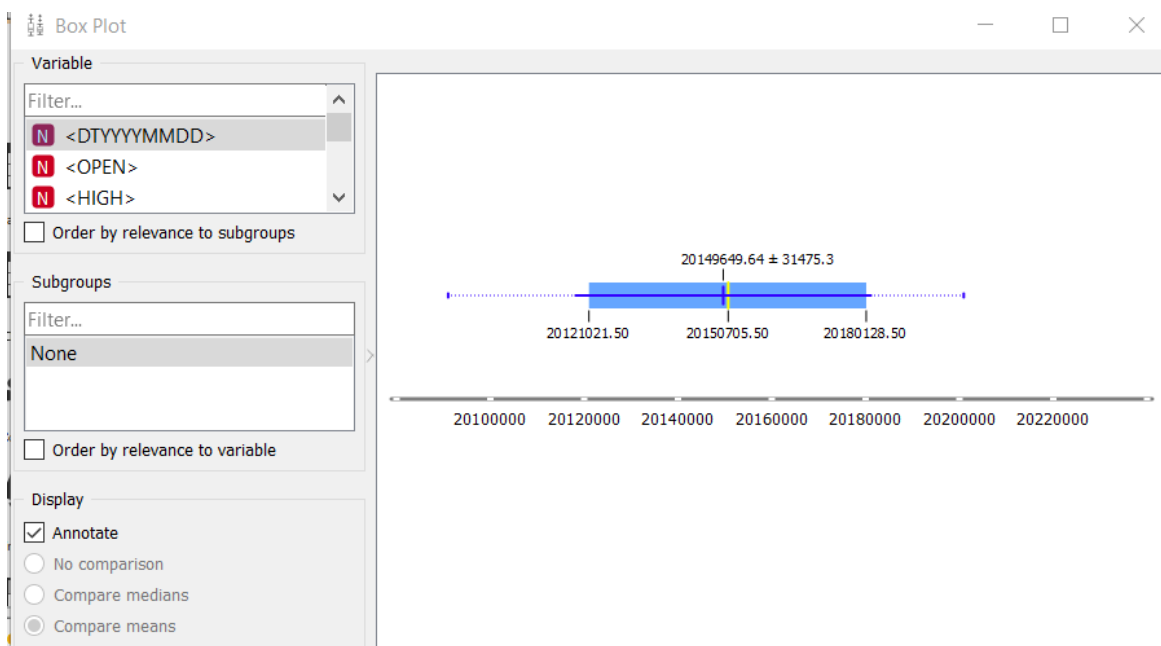
در شکل ۴ مشاهده می‌کنید که یک نمودار Scatter نمایش داده شده است که در محور X تاریخ و در محور Y حجم معاملات نمایش داده شده است که بیانگر این است با گذشت زمان حجم معاملات نیز افزایش پیدا کرده است.



شکل ۵

همانطور که در شکل ۵ مشاهده میکنید با استفاده نمودار سه بعدی scatter سه ویژگی تاریخ در محور x ، قیمت باز شدن در محور y و حجم معاملات در محور z را نمایش داده ایم. در این نمودار میتوانیم بفهمیم که با گذشت زمان هم قیمت باز شدن و هم حجم معاملات افزایش پیدا کرده است.

نرم افزار orange



شکل ۶

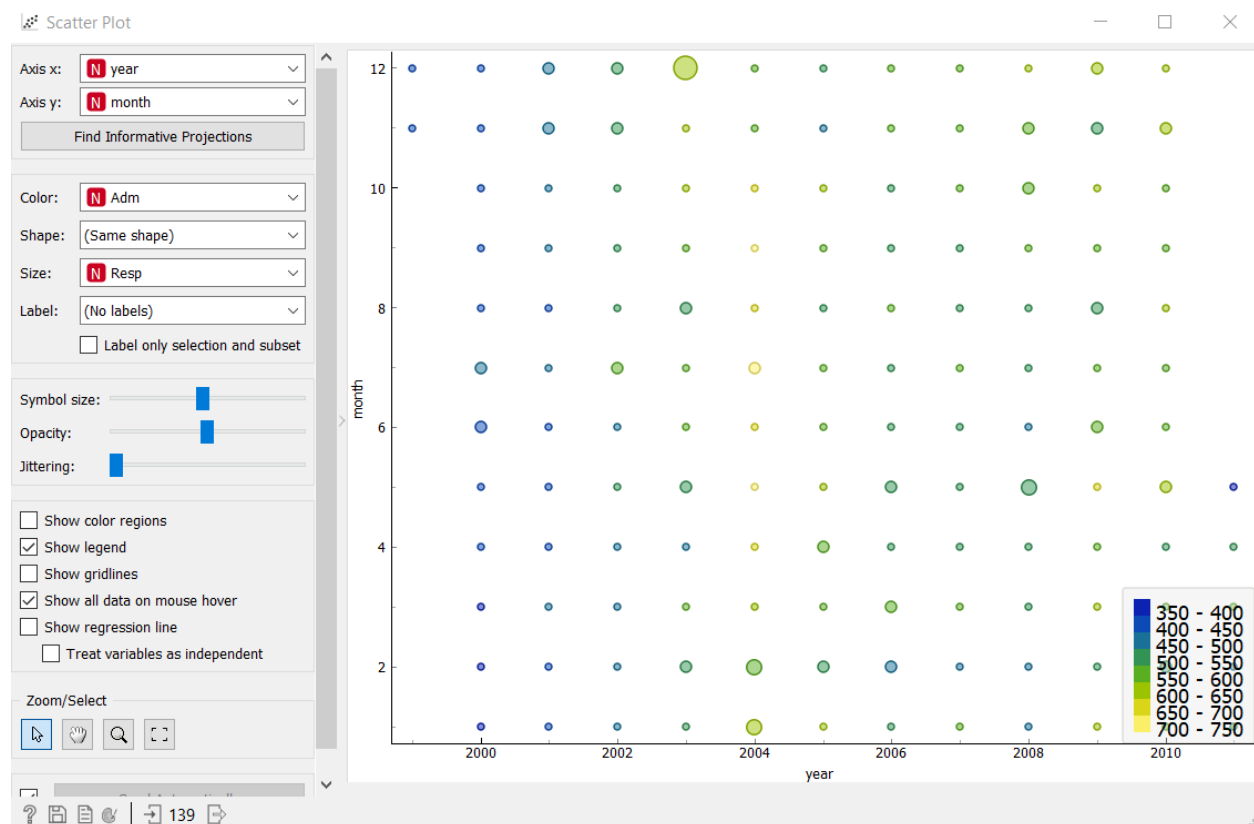
در این نرم افزار قابلیت نمایش نمودار های بیشتر وجود دارد.

همانطور که در شکل ۶ مشاهده میفرمایید نمودار box بر روی ویژگی تاریخ همراه با چارک ها نمایش داده شده است.



شکل ۷

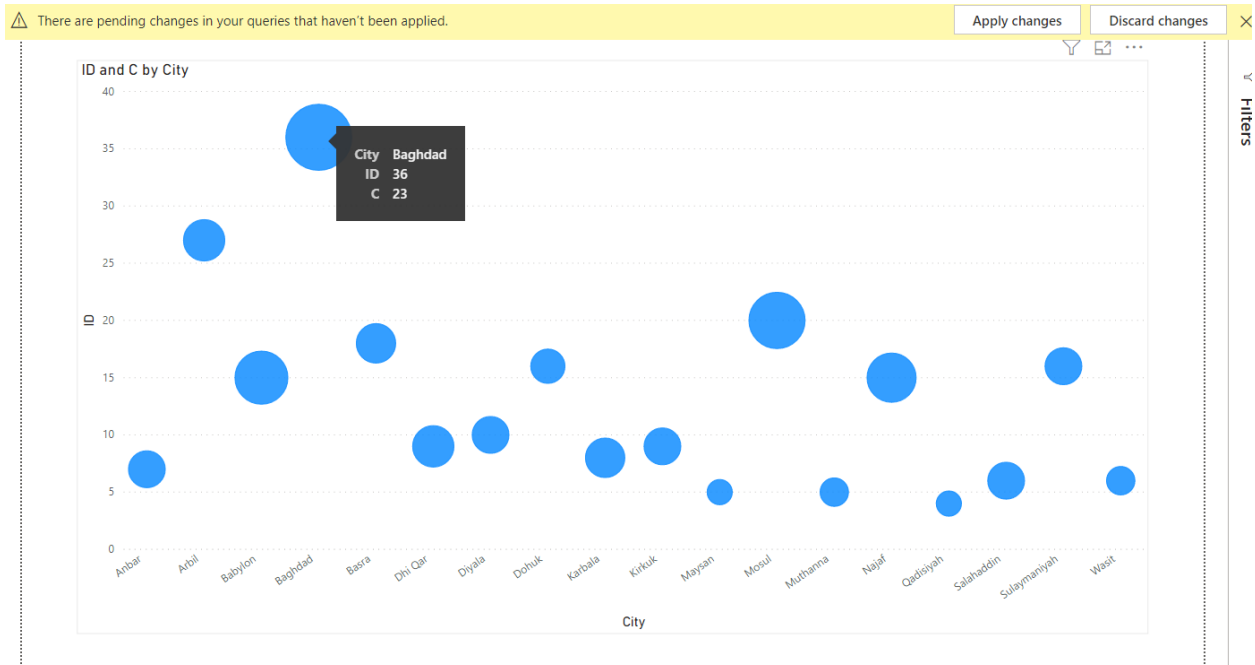
نکته قابل توجه در شکل ۷ میتوان به یک outlier اشاره کرد که باعث ایجاد بی نظمی و خطا در تحلیل شده است.



شکل ۸

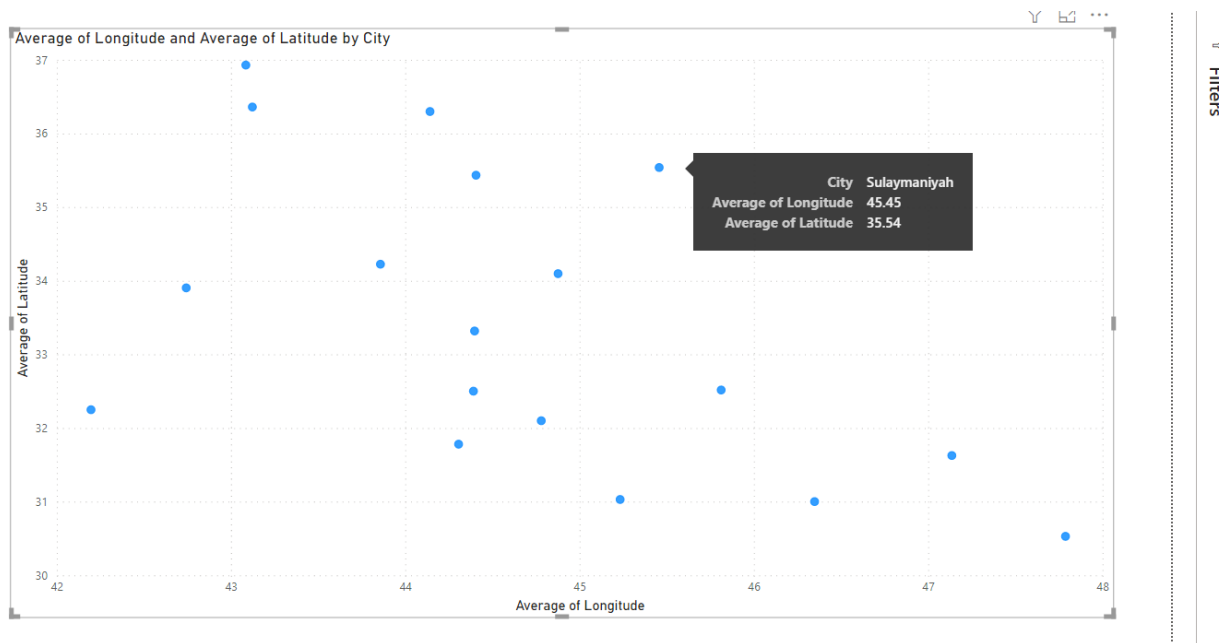
با استفاده از نرم افزار orange میتوان scatter plots را هم نمایش داد (شکل ۸) که در اینجا محور X سال، محور Y ماه، تغییرات رنگ adm و سایز اشکال را resp نمایش میدهد. در این شکل مشاهده میشود در ماه ۱۲ م سال ۲۰۰۳ بیشترین میزان ایست تنفسی در اورژانس اتفاق افتاده است.

نرم افزار Power BI



شکل ۹

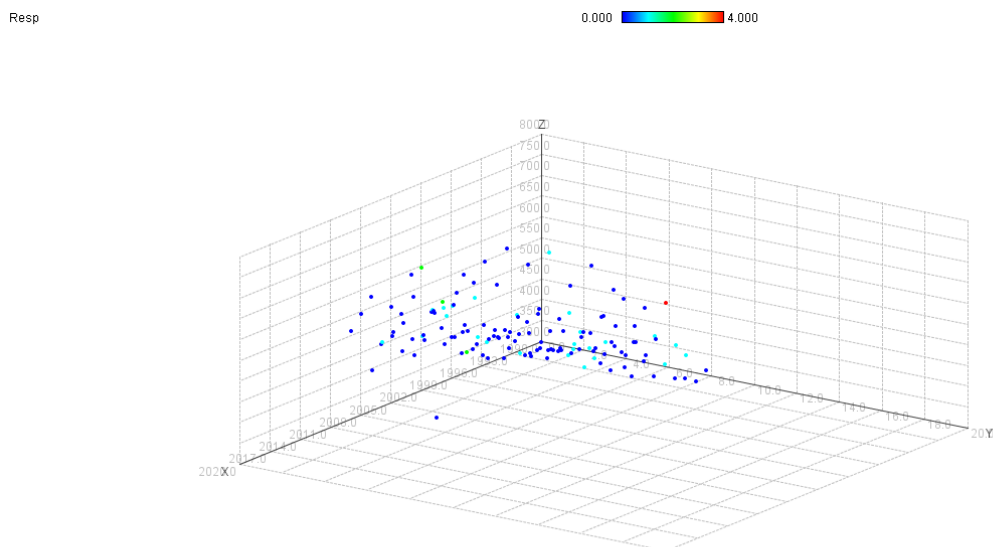
در این نرم افزار با استفاده از نمودار scatter بر روی مجموعه داده توریست های عراق نشان داده ایم (شکل ۹) که در هر شهر چه میزان توریست در مجموع دارد و چه تعداد از آنها در زمینه فرهنگی توریست هستند. برای مثال شهر بغداد ۳۶ توریست دارد که از این بین ۲۳ توریست مشغول کار فرهنگی میباشند.



شکل ۱۰

در شکل ۱۰ نشان در محور X طول جغرافیایی و در محور Y عرض جغرافیایی را قرار داده ایم به این صورت که مختصات هر شهر را بر روی نمودار به ما نشان میدهد. برای مثال شهر صلاح الدین در مختصات (۴۳,۴۵) و (۳۵,۵۴) قرار دارد.

نرم افزار Rapid Miner



شکل ۱۱

در این نرم افزار با استفاده از نمودار scatter 3D color یک نمودار ۴ بعدی رسم کرده ایم که در محور X سال، محور Y ماه، محور Z adm و تغییرات رنگ براساس Resp می باشد. که میتوان بررسی کرد در چه زمان هایی بیماران بخش اورژانس ایست تنفسی کرده اند.

همچنین در این نرم افزار میتوان انواع نمودارها از جمله Parallel، scatter matrix، Histogram و غیره را رسم کرد.

منابع

- [1] <https://archive.ics.uci.edu/ml/datasets/%3A+Simulated+Data+set+of+Iraqi+tourism+places#>
- [2] <https://archive.ics.uci.edu/ml/datasets/Horton+General+Hospital>
- [3] <https://academy.rapidminer.com/learning-paths/get-started-with-rapidminer-and-machine-learning>
- [4] <https://docs.microsoft.com/en-us/learn/powerplatform/power-bi>
- [5] <https://www.cs.waikato.ac.nz/ml/weka/>

[6] <https://orange.biolab.si/>