

به نام خدا

سینا براری ۹۹۳۲۱۲۰۱۲

تمرین اول

- نمودار خطی (Line chart)

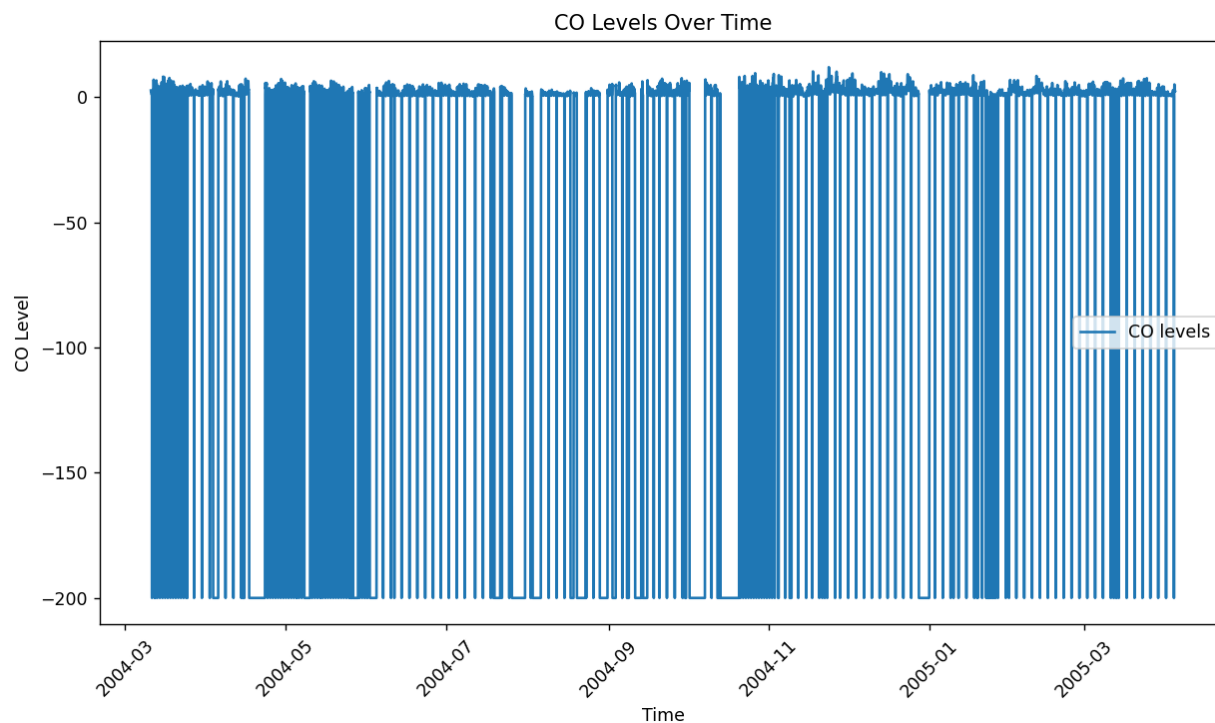
برای این نمودار از مخزن داده های UCI Repository ، از مجموعه داده Air Quality استفاده کردم.

لینک دیتاست : <http://archive.ics.uci.edu/dataset/360/air+quality>

این مجموعه داده شامل اندازه گیری های مختلفی از کیفیت هوا در یک شهر ایتالیایی است. این مجموعه شامل بیش از ۹۰۰۰ نمونه میباشد و ۱۵ ویژگی دارد و همچنین داده ها به صورت رگرسیون (پیوسته) می باشند.

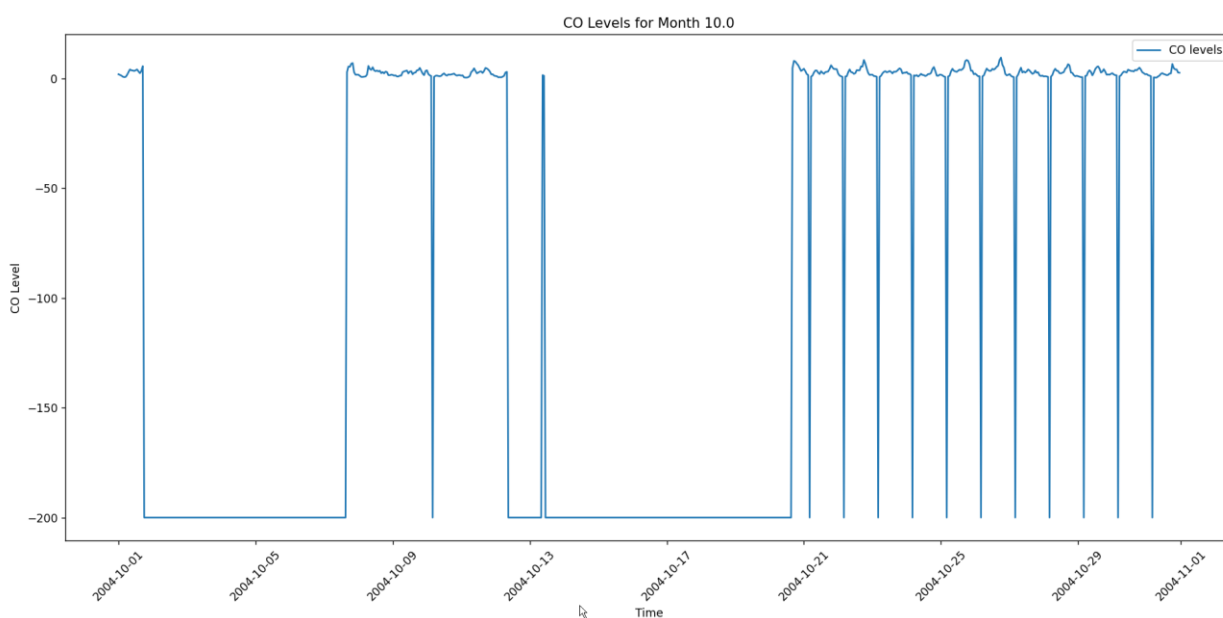
در این مجموعه داده به دلیل اینکه دارای دیتا های زیادی هستیم و در بازه زمانی کوتاه به نسبت تعداد داده ها که از ماه ۳ سال ۲۰۰۴ تا ماه ۴ سال ۲۰۰۵ داریم و اینگونه که روند تغییرات را هر دو ماه و کلی نشان بدهیم مناسب نیست و نمودار خوانایی مناسب را ندارد.

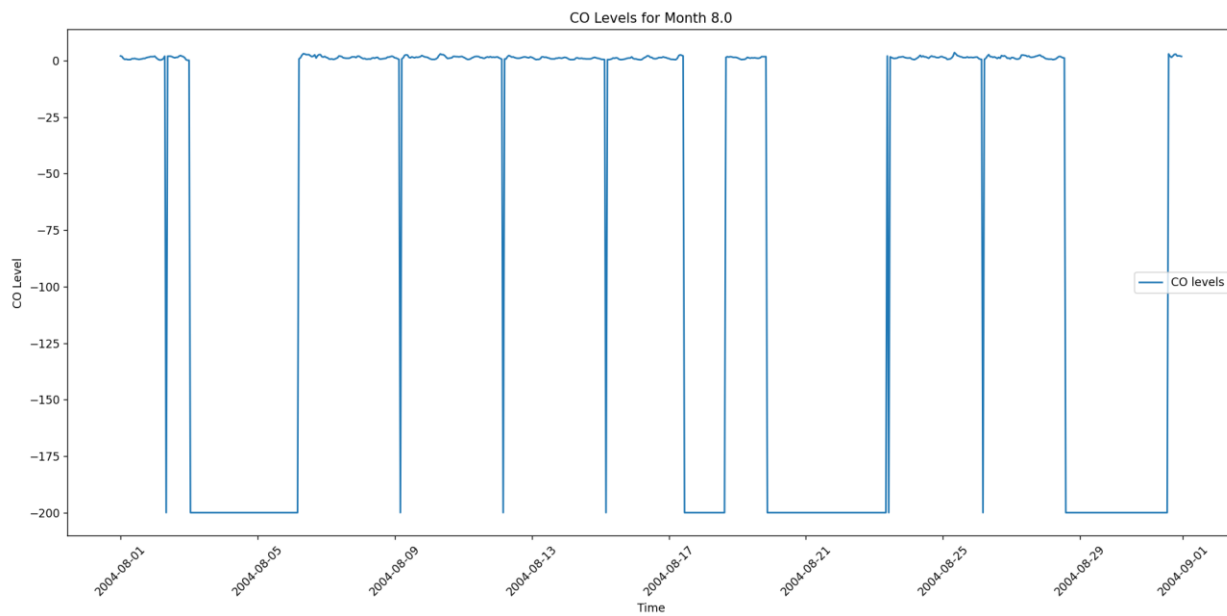
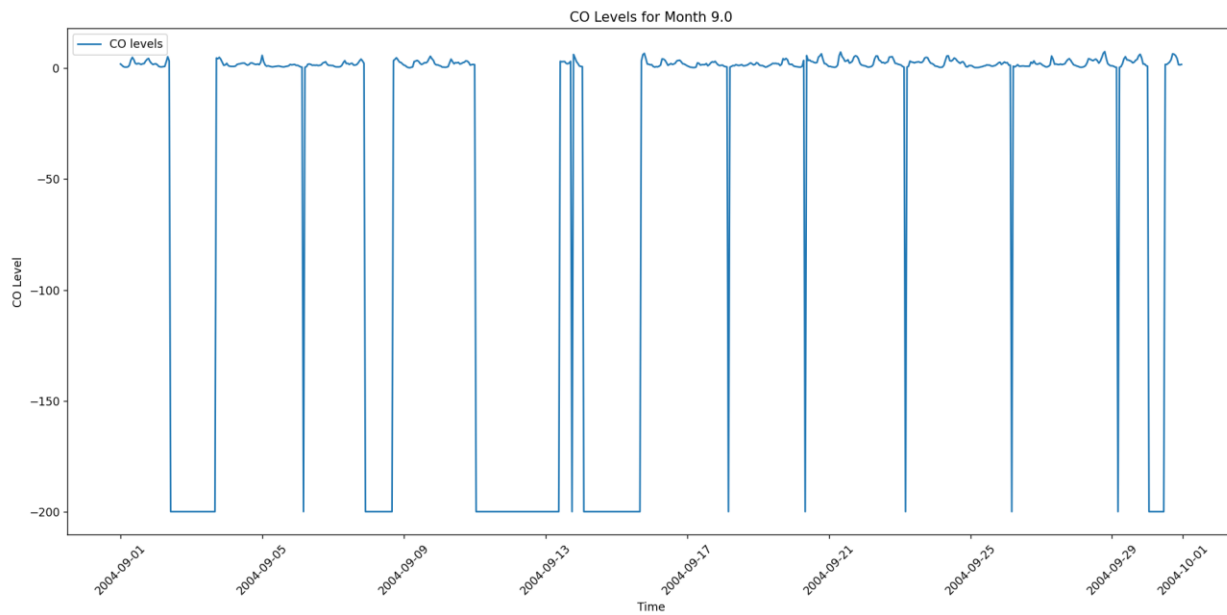
شکل نمودار به صورت زیر شد.

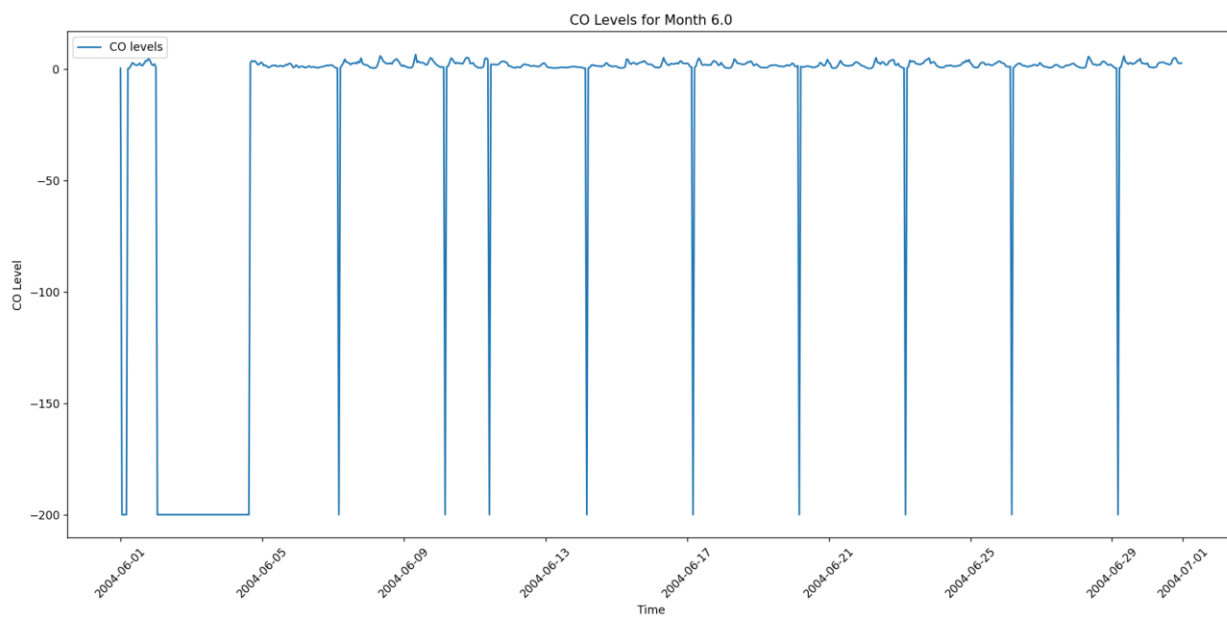
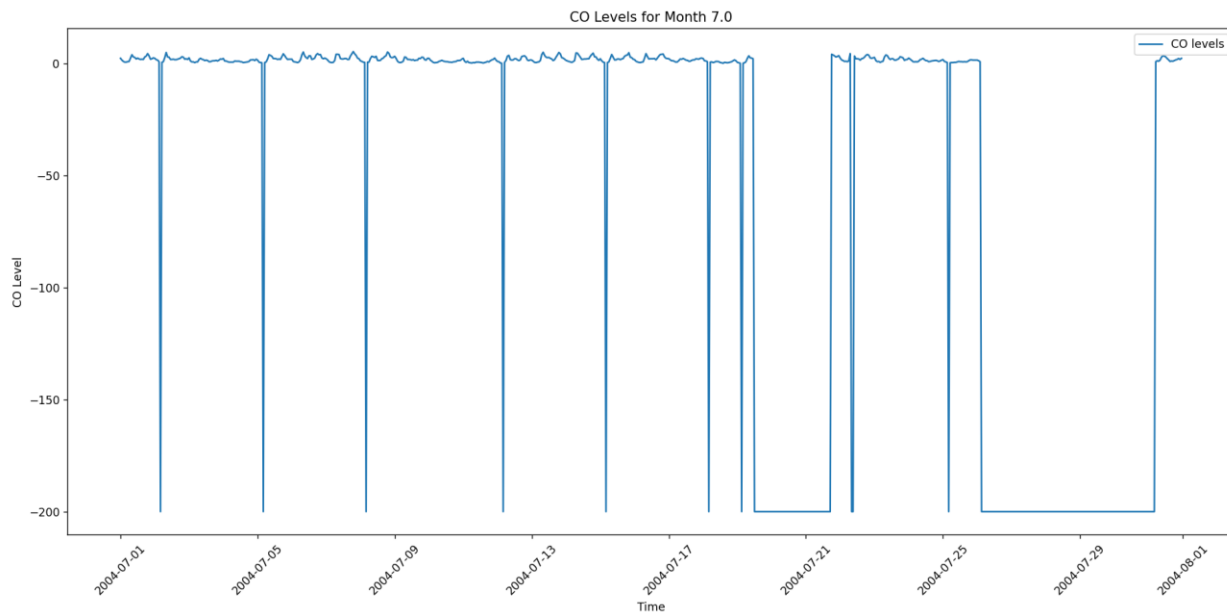


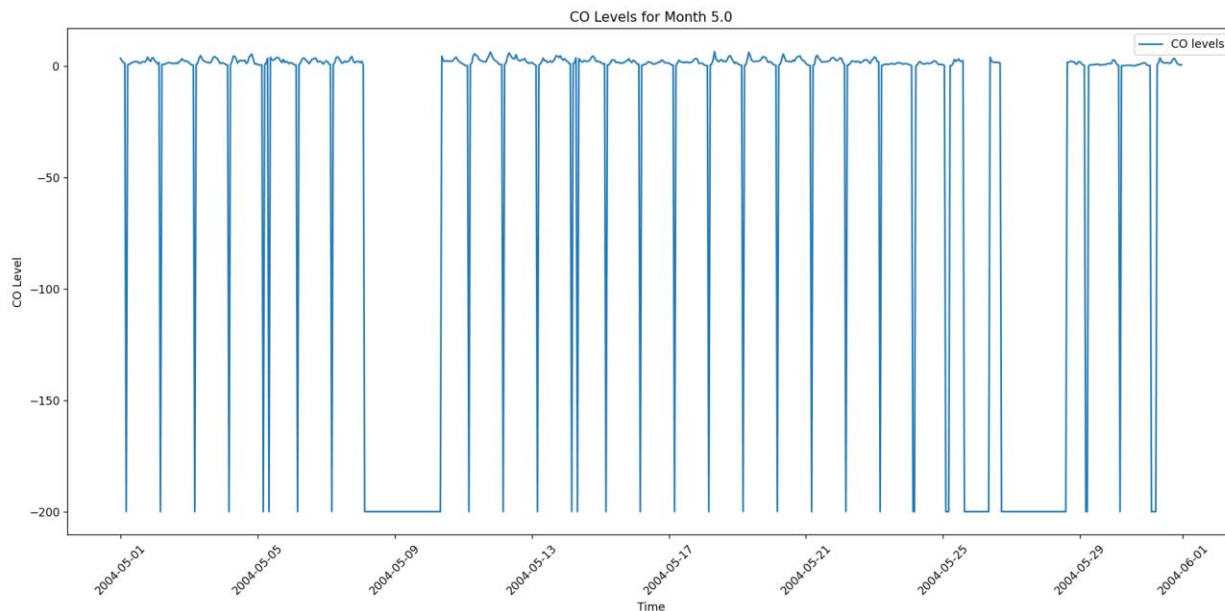
همانطور که مشاهده میکنیم بالا و پایین و شیب تغییرات بسیار زیاد است و در کل برای تحلیل وضعیت و کیفیت هوایی یک شهر مناسب نیست برای همین این داده ها را به صورت چند نمودار بررسی میکنیم.

حال نمودار برخی ماه ها را نمایش میدهم که به صورت زیر است :









همانطور که مشاهده میکنید جزئیات تغییرات بهتر قابل مشاهده است.
و همینطور میتوان به صورت هفتگی تقسیم کرد ولی باید از یکسری داده ها صرف نظر کرد.

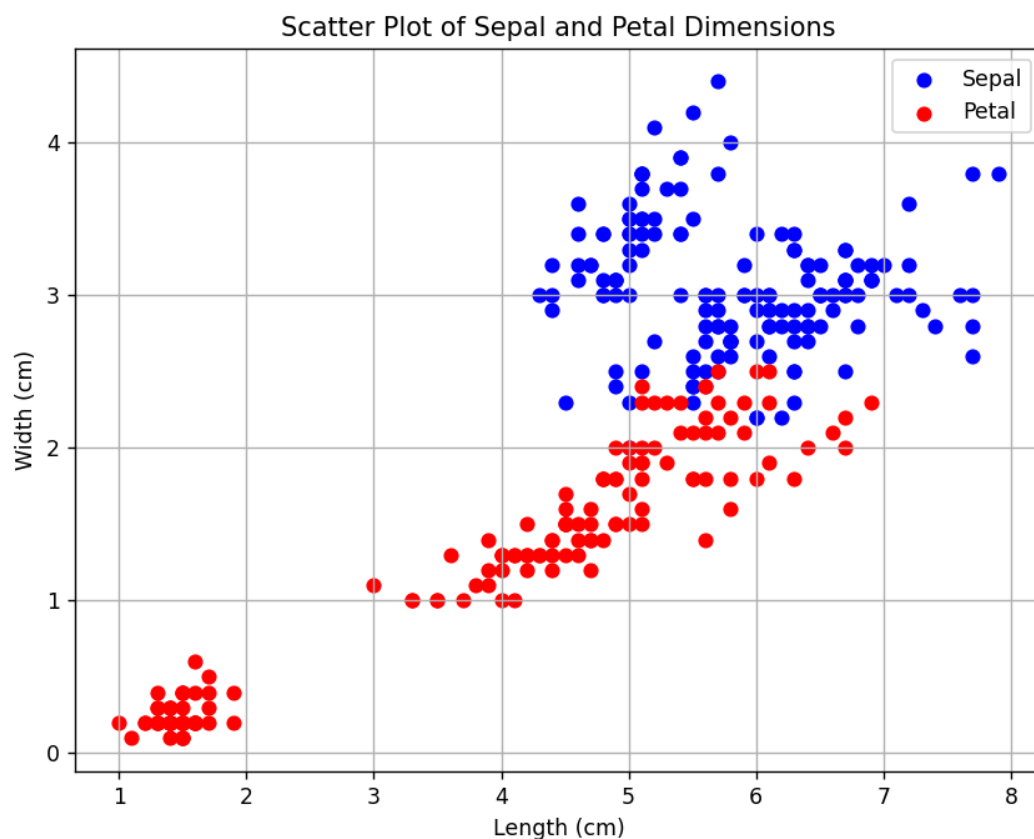
• نمودار پراکندگی (Scatter Plot)

برای رسم این نمودار از انباره UCI KDD Archive استفاده کردیم و لینک مجموعه داده به صورت زیر است.
لینک :

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

مجموعه داده Iris شامل اطلاعات در مورد گل های زنبق است و شامل چهار ویژگی (طول و عرض گلبرگ ها و طول و عرض کاسبرگ) است ، که می توان از آن ها برای رسم نمودار پراکندگی استفاده کرد.

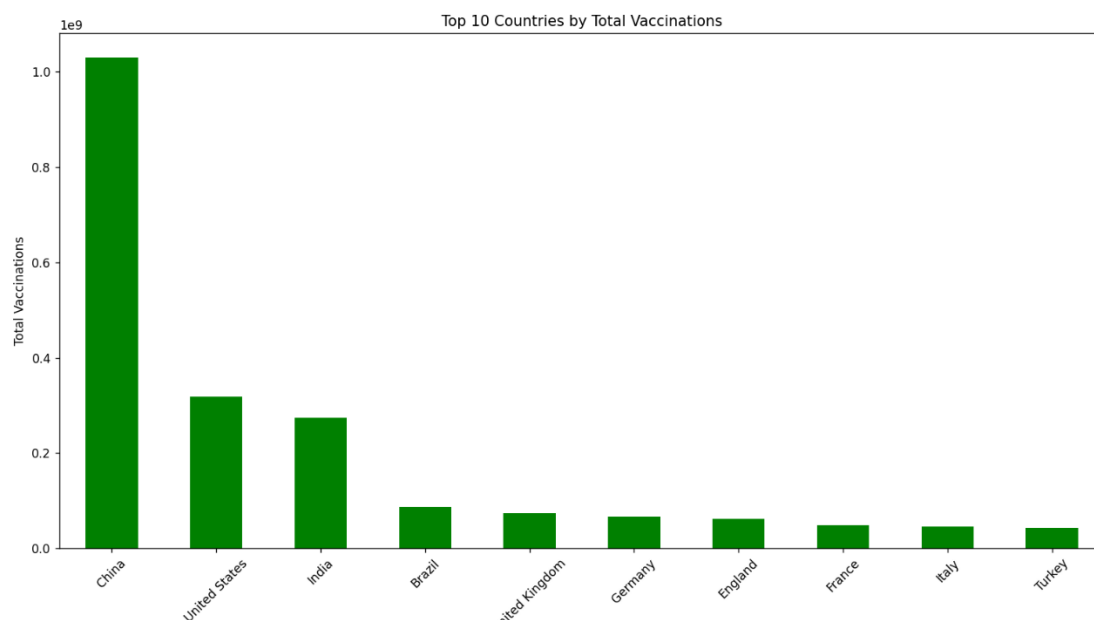
نمودار رسم شده به صورت زیر است.



- نمودار میله ای (Bar Chart)

برای رسم این نمودار از مخزن داده Kaggle استفاده کردیم و مجموعه داده COVID-19 World Vaccination Progress استفاده کردیم که در این نمودار ۱۰ کشوری که بیشترین واکسیناسیون را داشتن را نمایش میدهد. البته لازم به ذکر است برای سال ۲۰۲۱ میلادی است و دیتاها برای حال حاضر نیست.

نمودار به شکل زیر است :



لینک این دیتاست :

<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress/versions/129?resource=download>

• نمودار دایره ای (Pie Chart)

برای رسم این نمودار از انباره UCI Repository استفاده کردم و از مجموعه داده Iris استفاده کردم که مربوط به گل ها میباشد.

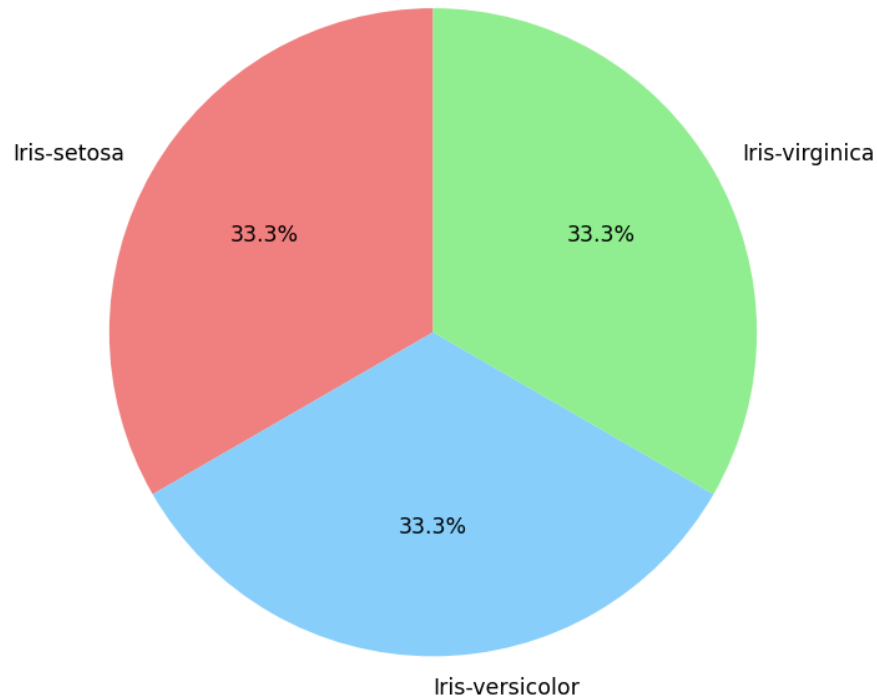
که در رسم نمودار پراکندگی هم از این دیتاست استفاده کردم.

لینک دیتاست :

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

و نمودار به شکل زیر میباشد.

Iris Flower Species Distribution (Pie Chart)



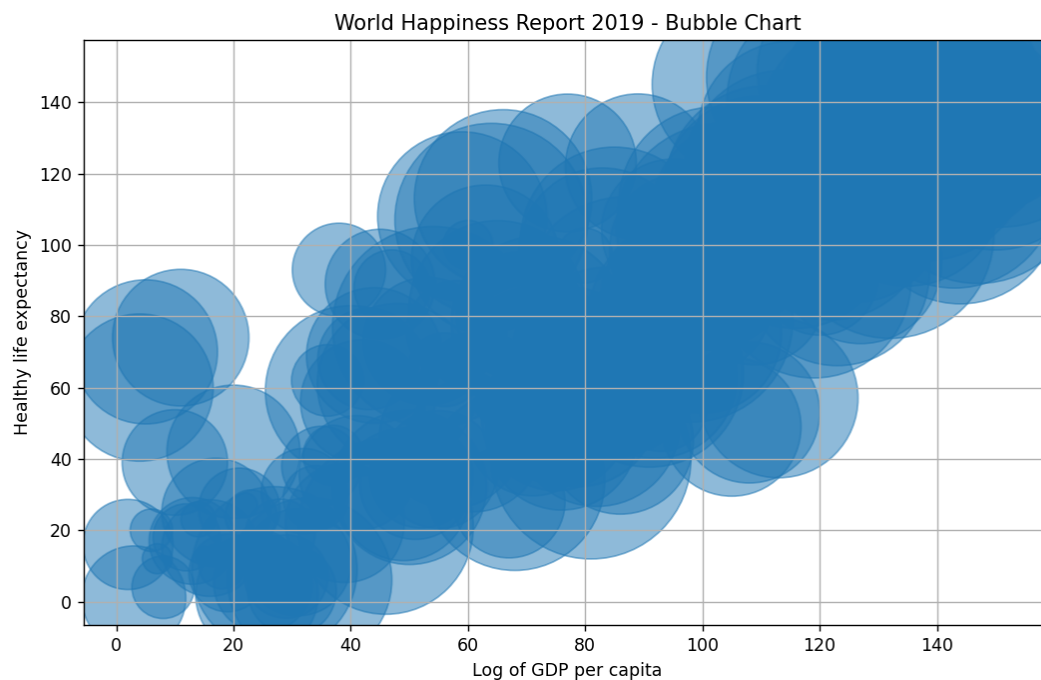
همانطور در نمودار میبینیم انواع گل ها و درصدی که دیتا ها به این ها اختصاص داده شد نمایش میدهیم.

• نمودار حبابی (Bubble Chart)

برای رسم این نمودار ما از مخزن داده Delve Datasets استفاده کردیم و از دیتاست World Happiness Report استفاده کردیم که در مورد میزان شادی و رضایت در کشور ها می باشد.
لینک مجموعه داده :

https://github.com/imgraham1/539_FinalProject/blob/master/world-happiness-report-2019.csv

نمودار ما به شکل زیر می باشد :



همانطور که می بینید مقدار GDP و سلامت در دنیا است.

• نمودار خرمن (Sankey Diagram) :

برای رسم این نمودار از انبار UCI Repository استفاده کردیم و از مجموعه داده Wine Quality استفاده کردیم و لینک آن به صورت زیر است :

<https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv>

و نمودار آن به شکل زیر است :



این نمودار ، تعداد کیفیت های خوب و بد وین را نمایش میدهد که به صورت و فرم نمودار خرمن ما آن را نمایش میدهیم.

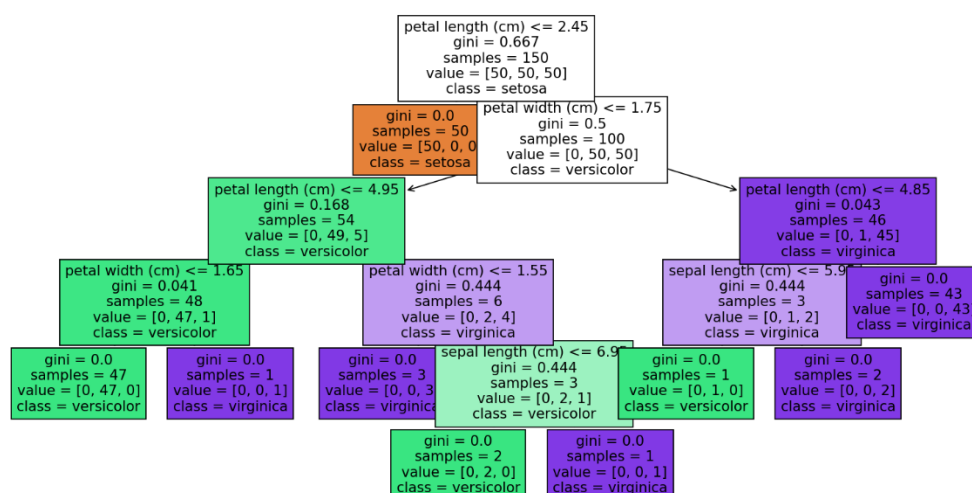
- نمودار درختی (TreeMap Chart) :

برای رسم این نمودار از ریبازیتوری **UCI Repository** استفاده کردیم و مجموعه داده **Iris** استفاده کردم که داده هایی از گلبرگ های سه نوع گل است که مقدار طول گلبرگ ، مقدار جینی و ... را نمایش میدهد.

لینک این مجموعه داده :

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

نمودار به این صورت می باشد :



همانطور که میبینید انواع گل ها و گلبرگ را به صورت درختی نمایش میدهد.

• نمودار هیستوگرام (Histogram Chart) :

برای رسم این نمودار از رپازیتوری **UCI Repository** استفاده کردیم و مجموعه داده

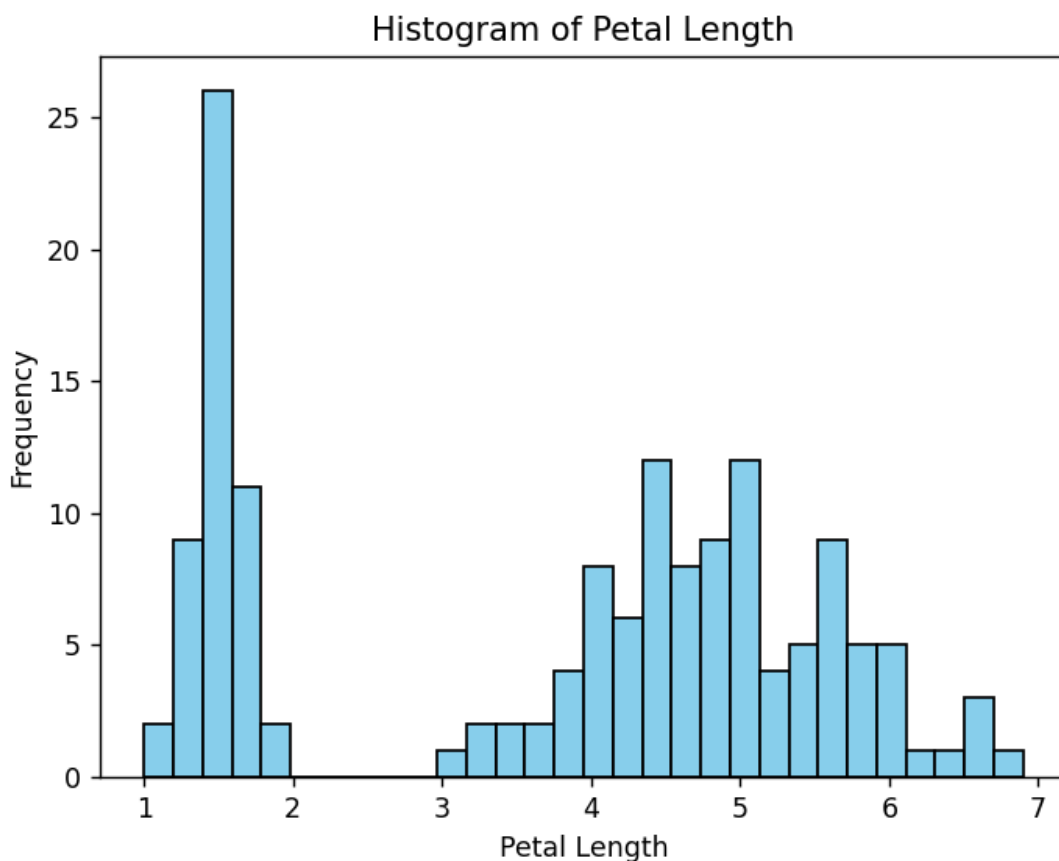
iris استفاده کردیم که پارامترهای مختلف از داده هایی از گلبرگ های انواع مشخصی از

گل ها است مانند مقدار طول گلبرگ ، مقدار جینی و

لینک این مجموعه داده :

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

نمودار به این صورت می باشد :



همانطور که میبینیم این نمودار توزیع اندازه طول گلبرگ ها را براساس تعداد و تکرار این ها به نمایش گذاشته.

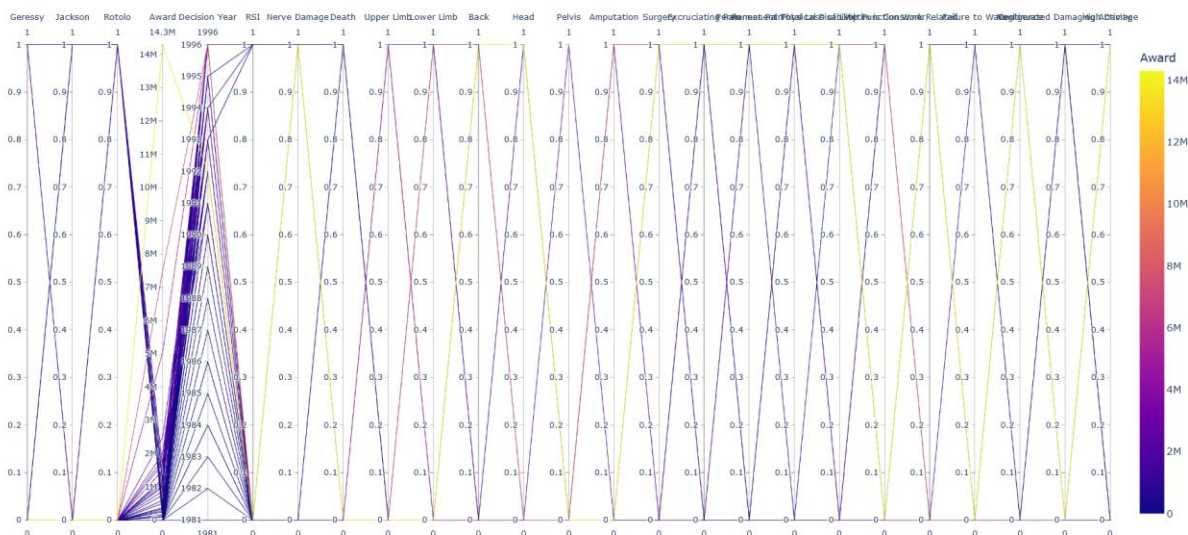
- نمودار هماهنگی موازی (Parallel Coordinates) :

برای رسم این نمودار از ریپازیتوری **Statlib** استفاده کردیم و مجموعه داده **LPR** استفاده کردم که این مجموعه داده برای تعیین حواله توسط قاضی واینسین انجام شده که بررسی میکند آیا غرامت خواسته شده با مقدار آسیبی که توسط متهمان وارد شده یکسان است یا خیر.

لینک این مجموعه داده :

<http://lib.stat.cmu.edu/datasets/lpr.zip>

نمودار به این صورت می باشد :



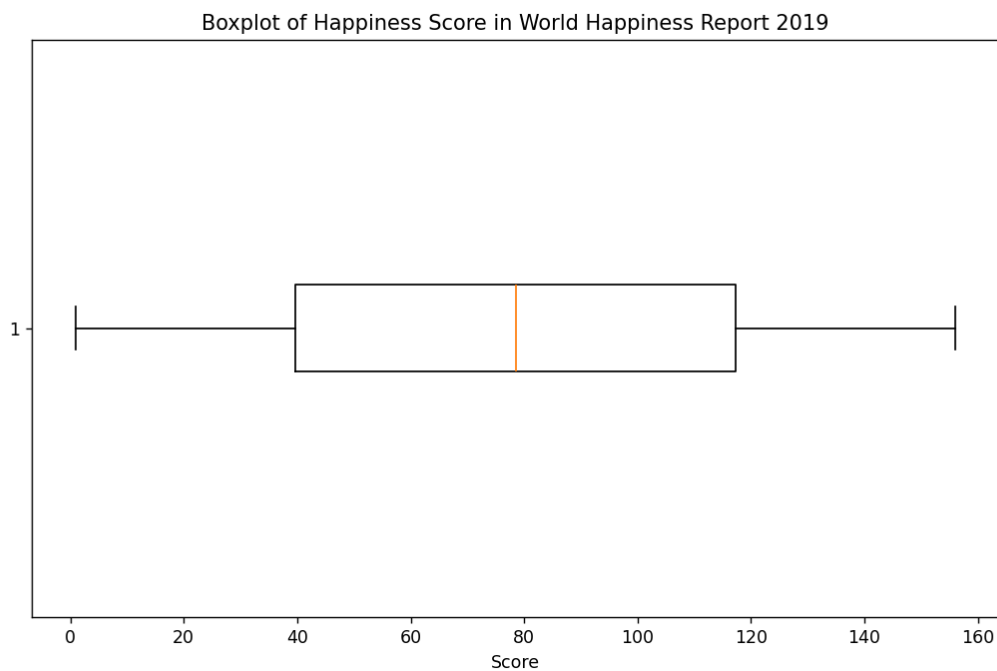
همانطور که مشاهده میکنید در این دیتاست این نمودار برای feature هایش مناسب نیست چون مقدارشان بیشتر ۰ و ۱ است و داده ها به یک بخش نمی‌رسند. و پی می‌بریم ویژگی چهارم مناسب و کمک کننده است.

• نمودار جعبه‌ای (Box Plot) :

برای رسم این نمودار از رپازیتوری Delve Datasets استفاده کردیم و مجموعه داده World Happiness Report استفاده کردم که میزان شادی در کشور هاست و همچنین فاکتور هایی که بر روی شادی تاثیر گذارند در کشور ها چه مقدار است. لینک این مجموعه داده :

https://github.com/imgraham1/539_FinalProject/blob/master/world-happiness-report-2019.csv

نمودار به این صورت می‌باشد :



همانطور که میبینید مقداری که به کشور اختصاص داده شده و رتبه بندی شده اند میانه آن ها ۸۰ است و از ۱ تا ۱۶۰ است و چارک بندی آن را مشاهده میکنیم.

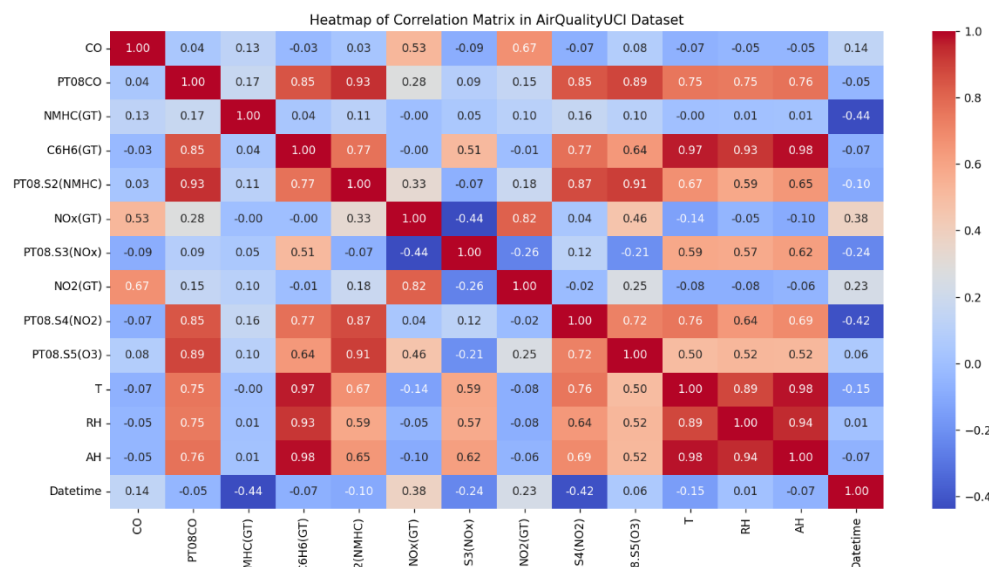
- نمودار حرارتی (Heatmap) :

برای رسم این نمودار از ریپازیتوری UCI Repository استفاده کردم و مجموعه داده Air Quality استفاده کردم که شاخص کیفیت هوای یک شهر ایتالیایی است.

لینک این مجموعه داده :

<http://archive.ics.uci.edu/dataset/360/air+quality>

نمودار به این صورت می باشد :



همانطور که در نمودار مشاهده می کنید کیفیت هوا را طبق پارامتر های مختلف درجه بندی کرده و از آب پرنگ به کم رنگ تا تقریبا سفید و تا قرمز پر رنگ می رود که اختلاف را به وضوح در جاهای مختلف نمایان است.

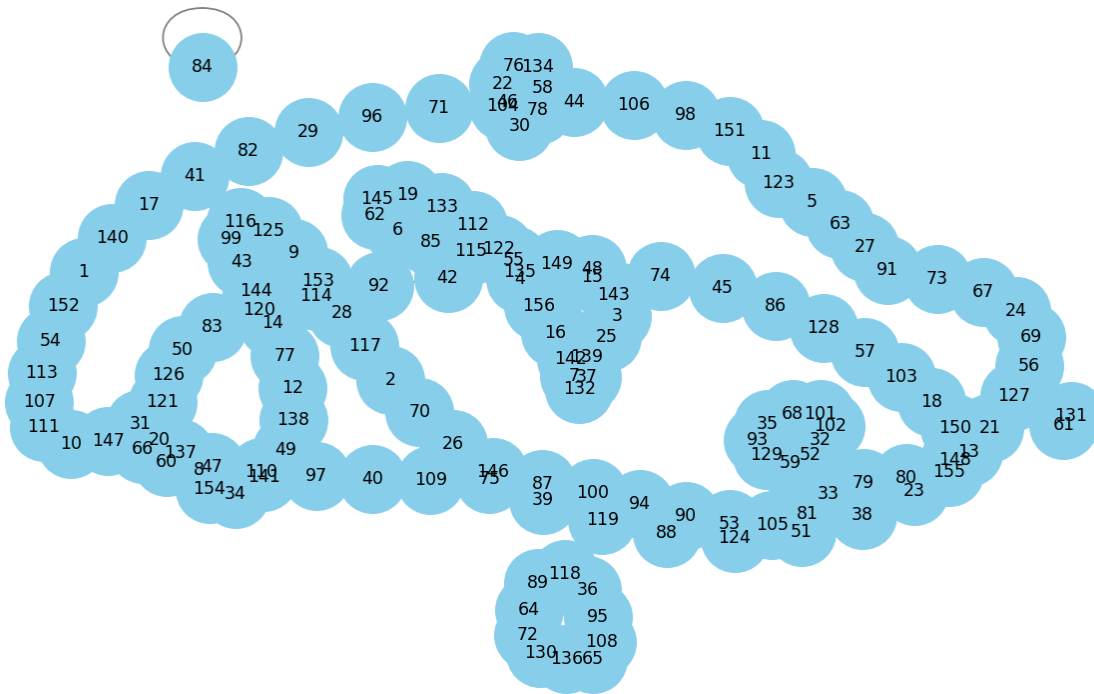
• نمودار گراف شبکه ای (Network Graph) :

برای رسم این نمودار از رپازیتوری Delve Datasets استفاده کردیم و مجموعه داده World Happiness Report استفاده کردیم که میزان شادی در کشور هاست و همچنین فاکتور هایی که بر روی شادی تاثیر گذارند در کشور ها چه مقدار است.

لینک این مجموعه داده :

https://github.com/imgraham1/539_FinalProject/blob/master/world-happiness-report-2019.csv

نمودار به این صورت می باشد :



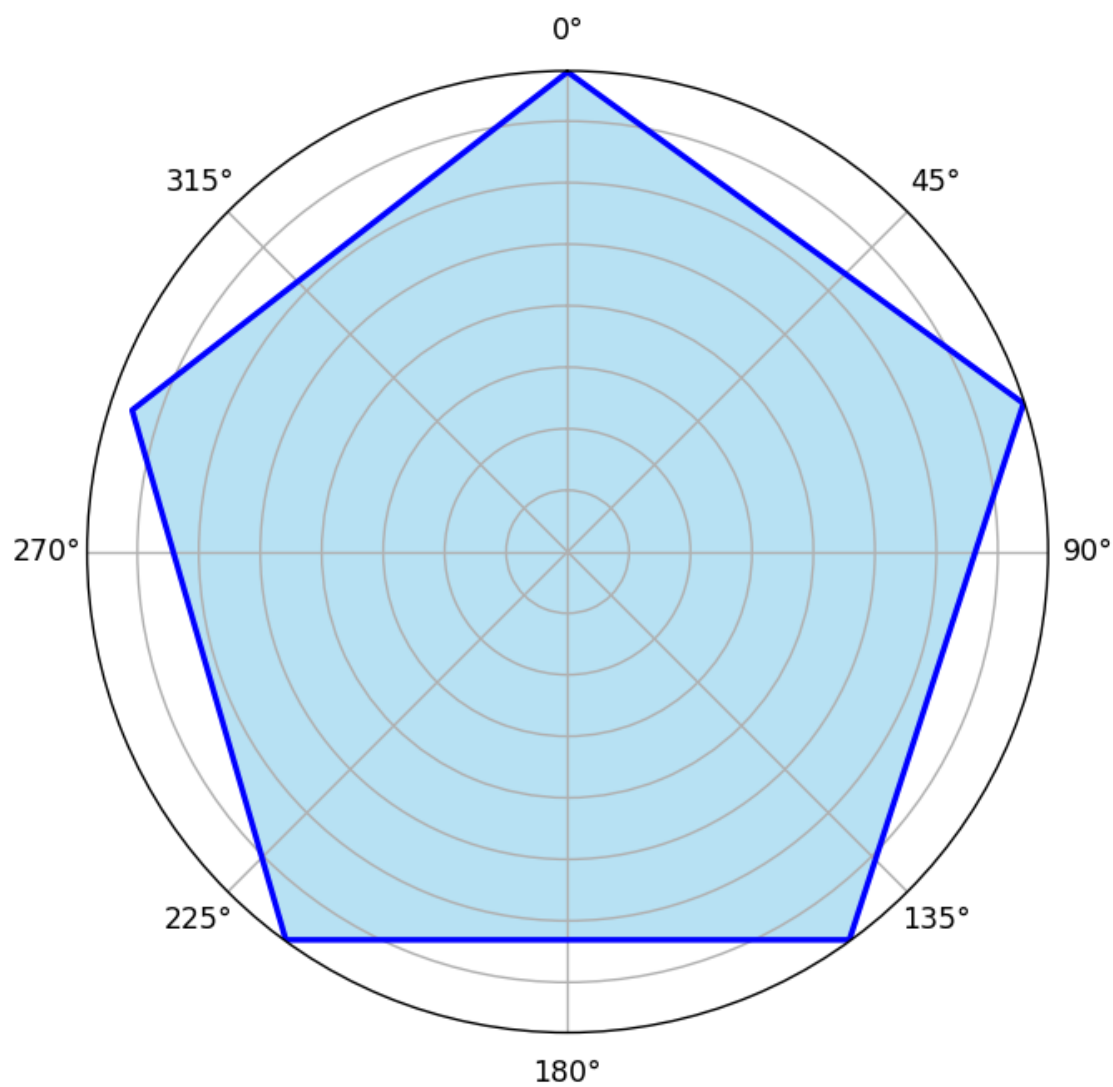
• نمودار راداری (Radar Chart) :

برای رسم این نمودار از ریبازیتوری Delve Datasets استفاده کردیم و مجموعه داده World Happiness Report استفاده کردیم که میزان شادی در کشورهاست و همچنین فاکتورهایی که بر روی شادی تاثیر گذارند در کشورها چه مقدار است.

لینک این مجموعه داده :

https://github.com/imgraham1/539_FinalProject/blob/master/world-happiness-report-2019.csv

نمودار به این صورت می باشد :



با تشکر از همراهی شما