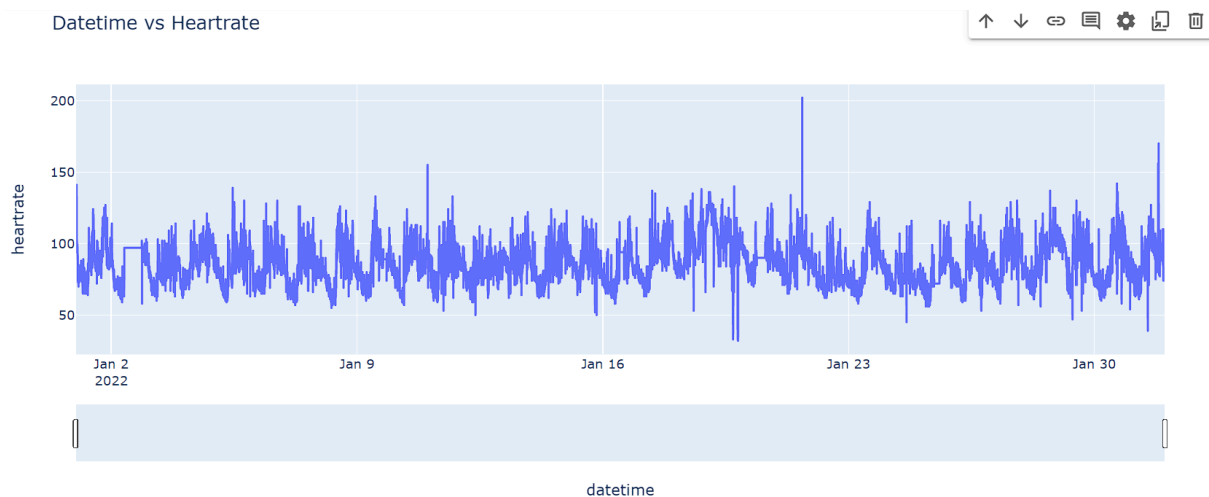


# Heart Rate anomaly detection

## Isolation Forest:

- Isolation Forest is a machine learning algorithm that efficiently identifies anomalies in data by creating isolation trees.
- These trees isolate anomalies by randomly selecting features and split values, recursively partitioning the data.
- Anomalies, being less common, are expected to be isolated faster, resulting in shorter paths in the tree.

## Heart Rate Vs DateTime:



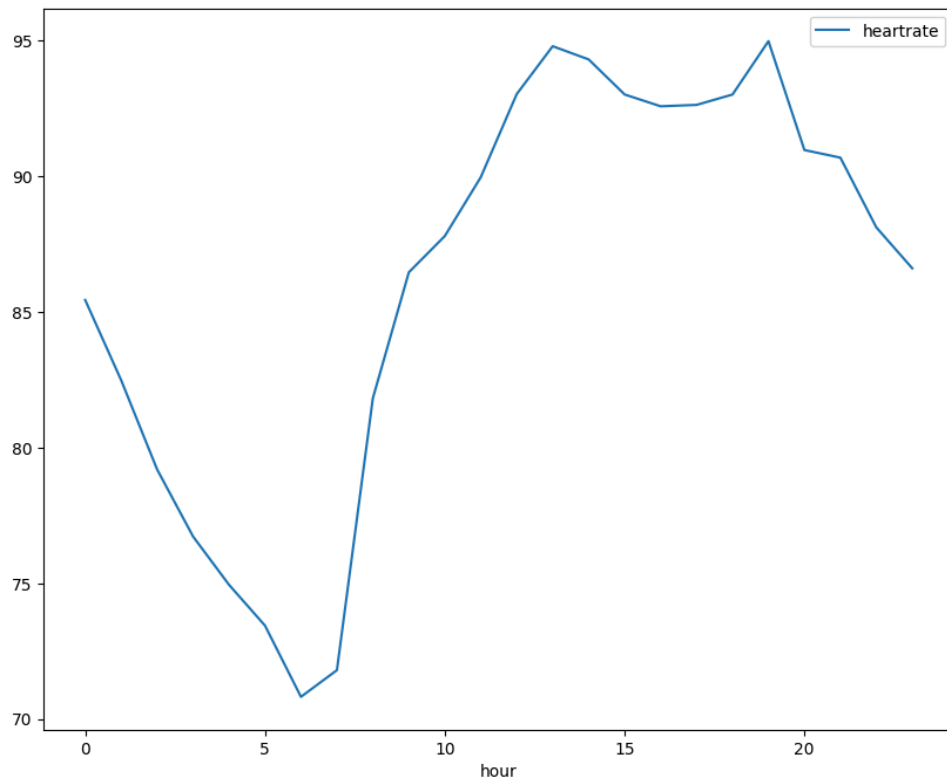
We can clearly see some outliers (drops and spikes) in the data. Our task is to detect those outliers along with contextual outliers which don't seem to be outliers but are.

## Mean heart rate per hour of the day:

From the below given graph, we can see plot of mean heart rate every hour of the day.

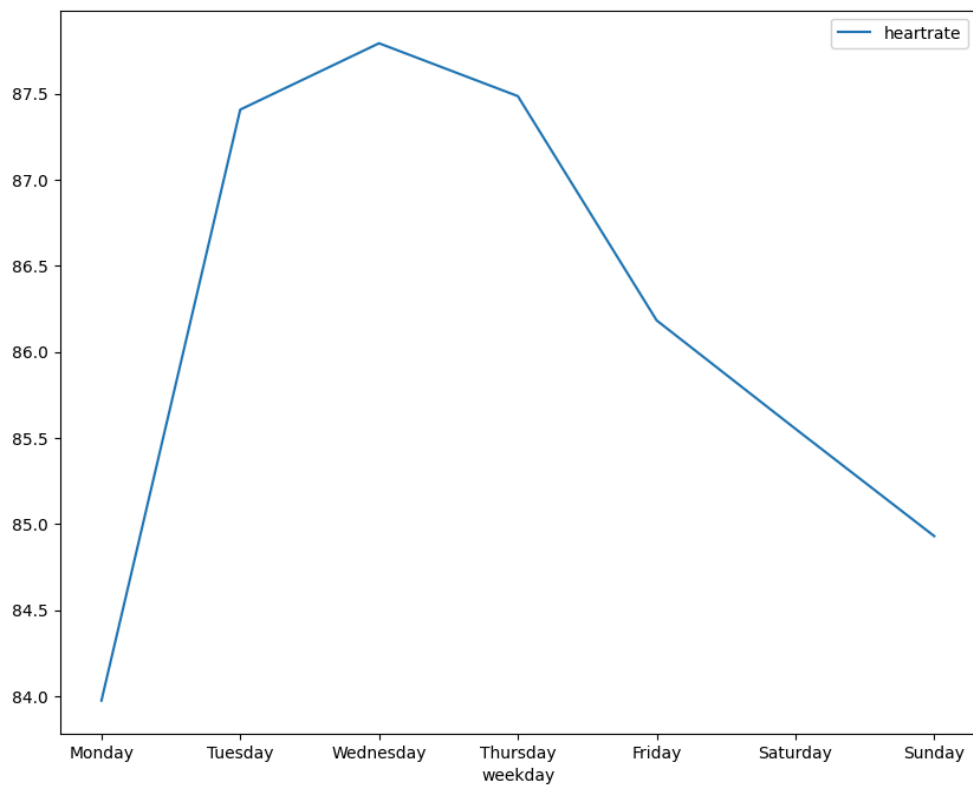
Heart rate is highest between 10 AM to 8 PM. It may be due to work activity or physical activity.

Heart rate is on lesser side from 8 PM to 8 AM because of sleep/resting period.

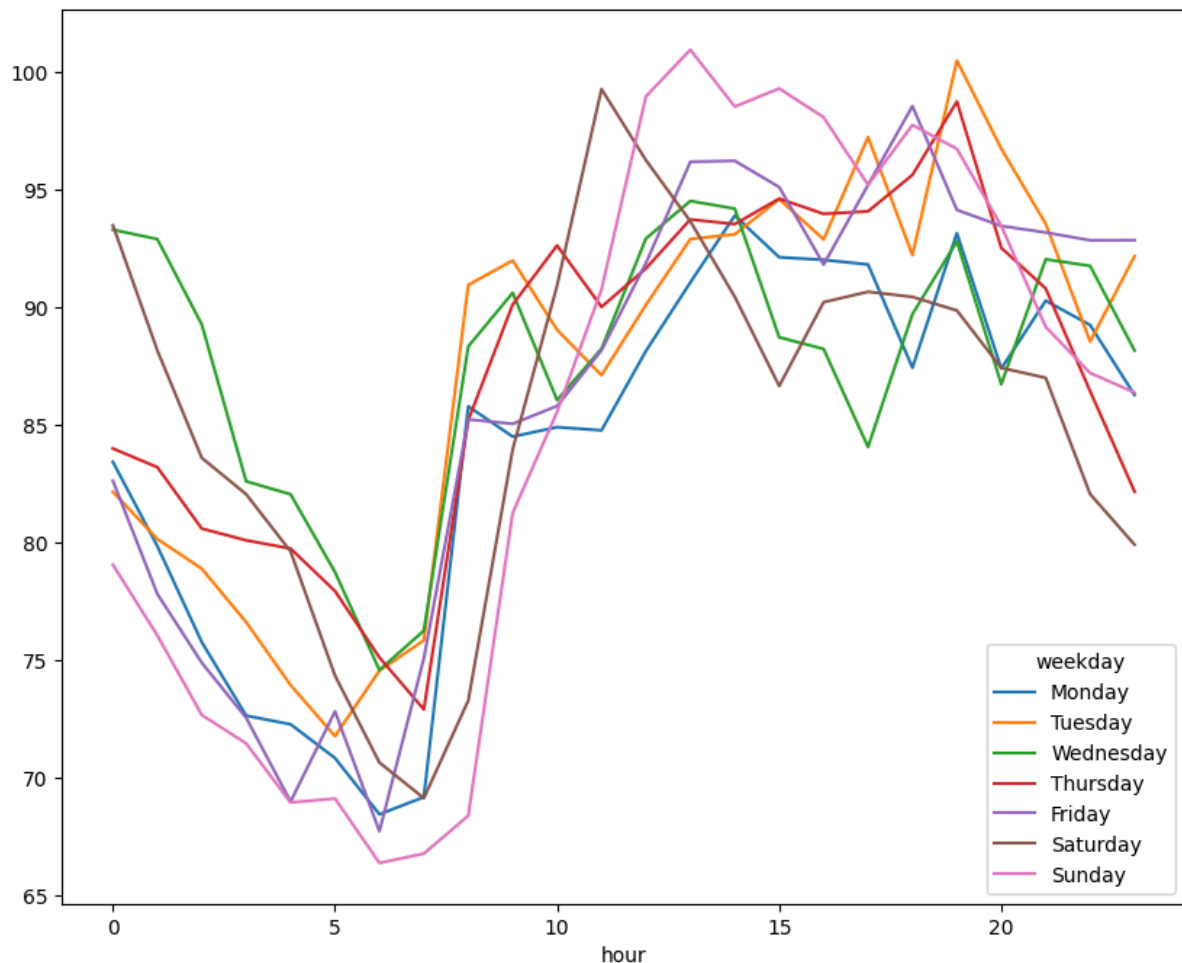


### Mean heart rate per day of the week:

Heart Rate is higher on weekdays like Tuesday, Wednesday, Thursday, while it's lower on weekends.



## Heart Rate hourly on every weekday:



## Insights from the data analysis and visualisation:

Heart rate is depending on the hour of the day so we need to consider that factor to catch contextual anomalies.

It is also dependent on the day of the week, so we will also consider that as a factor.

The mean heart rate of the person is 86 which is good.

## Action:

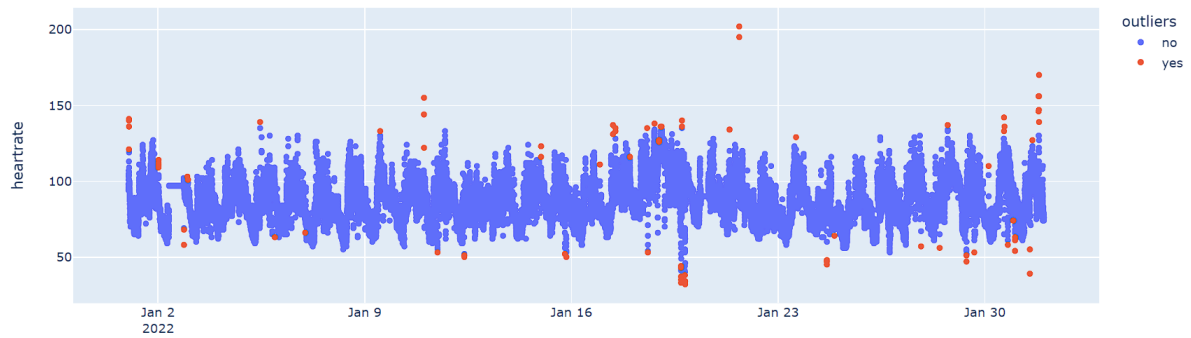
Created a new dataframe with three features: **Heart Rate**, **Day**, **Hour**.

Used Isolation Forest model to find anomalies. Considered 0.2% contamination.

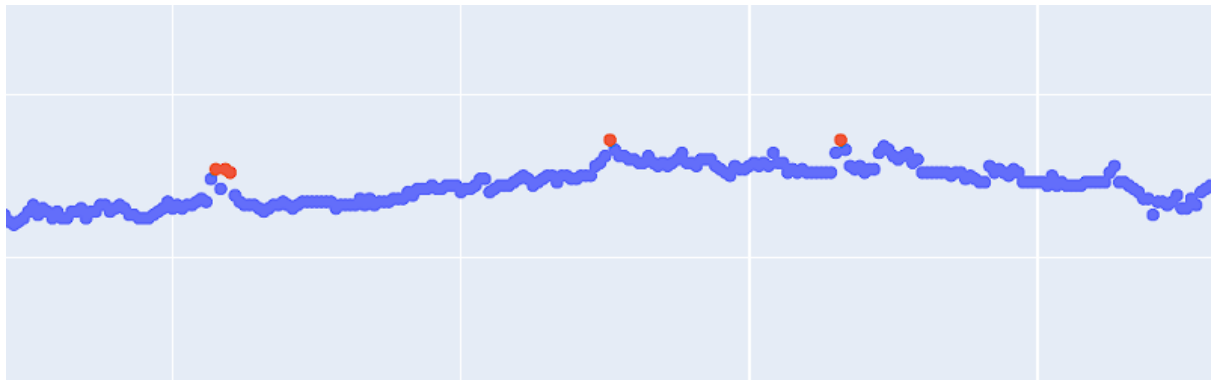
## Results:

Total count of detected anomalies is 89. They are coloured in red in below attached plot.

We can clearly see spikes and downs are detected as anomalies which is pretty intuitive. Other than that, some points are detected as anomalies which doesn't seem outliers but are contextual outliers.



Those points which don't seem to be anomalies, if seen by zooming are clearly anomalies.



## Conclusion:

Our model has successfully detected anomalies.

Out of detected anomalies, 20 are between 12 AM to 8AM while remaining others are between 8 AM to 8 PM.

Hence, the majority of anomalies are in day time.

Please refer to the colab notebook for complete code and analysis:

<https://colab.research.google.com/drive/1Rr6tY9smtSZRjyoEyRt-airYGTfKflqq#scrollTo=hJM6YtFGY5sR>