Mini-Project 1

ECE 471

Fall 2024

Cindy(hccheng3), Weber(powei2)

Task 1

Q1: Overview of the .csv files
Take data from folder 'clear-night' for example:

- ctl.csv

	ts	agent_id	throttle	steer	brake
0	22555	0	0.9	-0.013	0
1	22556	0	0.9	-0.0031	0
2	22557	0	0.9	-0.0102	0
3	22558	0	0.9	-0.0081	0
4	22559	0	0.9	-0.0001	0

- cvip.csv

	ts	agent_id	cvip	cvip_x	cvip_y	cvip_z
0	22555	0	500.491	198.767	-95.833	-499.82
1	22556	0	5.59558	195.567	-90.833	0.1
2	22557	0	5.59237	195.567	-90.833	0.09541
3	22558	0	5.58958	195.567	-90.833	0.08469
4	22559	0	5.58715	195.567	-90.833	0.06931

Task 1

Q1: Overview of the .csv files

Take data from folder 'clear-night' for example:

- traj.csv

	ts	agent_id	x	y	z	v
0	22555	0	192.362	-86.263	0.53933	0
1	22556	0	192.362	-86.263	0.49191	0
2	22557	0	192.362	-86.263	0.43837	0
3	22558	0	192.362	-86.263	0.37873	0
4	22559	0	192.362	-86.263	0.31298	0

Task 1

Q2:

a. The duration of the scene is: ______ frames (The ‘ts’ column is in the unit of frames).

Scene	**Duration**
clear-night	859 ts
clear-noon	752 ts
clear-sunset	756 ts
haze-noon	755 ts
haze-sunset	756 ts
rain-noon	401 ts

Task 1

Q2:

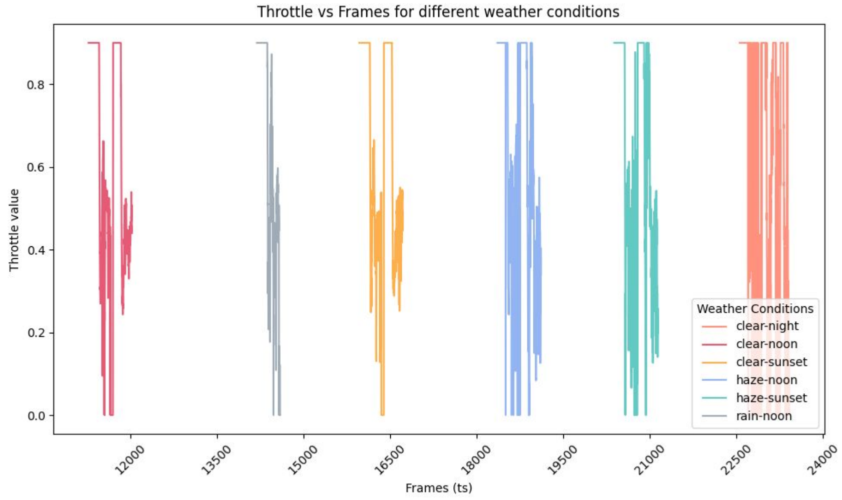
b. Mean and std of the features:

	clear-night	clear-noon	clear-sunset	haze-noon	haze-sunset	rain-noon
throttle	0.586 (0.357)	0.609 (0.282)	0.611 (0.281)	0.601 (0.298)	0.603 (0.298)	0.647 (0.272)
steer	0.004 (0.060)	0.001 (0.005)	0.000 (0.005)	0.001 (0.005)	0.001 (0.005)	-0.000 (0.003)
brake	0.138 (0.345)	0.061 (0.240)	0.057 (0.232)	0.060 (0.237)	0.052 (0.221)	0.037 (0.190)
cvip	33.425 (33.532)	19.533 (24.277)	20.055 (24.638)	20.209 (24.276)	20.144 (24.345)	7.050 (24.704)
x	191.533 (1.127)	192.920 (0.403)	192.946 (0.401)	192.941 (0.411)	192.942 (0.411)	192.631 (0.251)
y	-29.854 (39.593)	-31.341 (40.553)	-31.218 (40.408)	-31.827 (40.657)	-31.566 (40.665)	-63.877 (21.147)
v	6.074 (2.690)	6.927 (3.283)	6.884 (3.315)	6.891 (3.062)	6.887 (3.055)	6.357 (3.873)

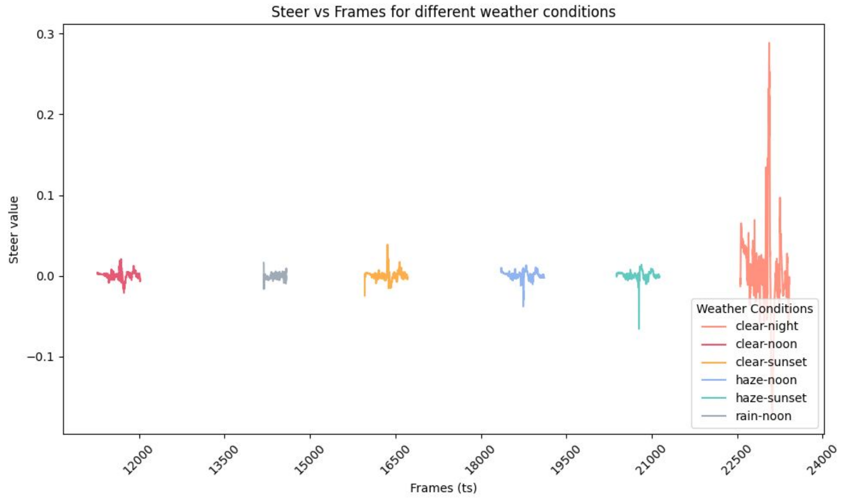
Task 1

Q3: Campaign result visualization

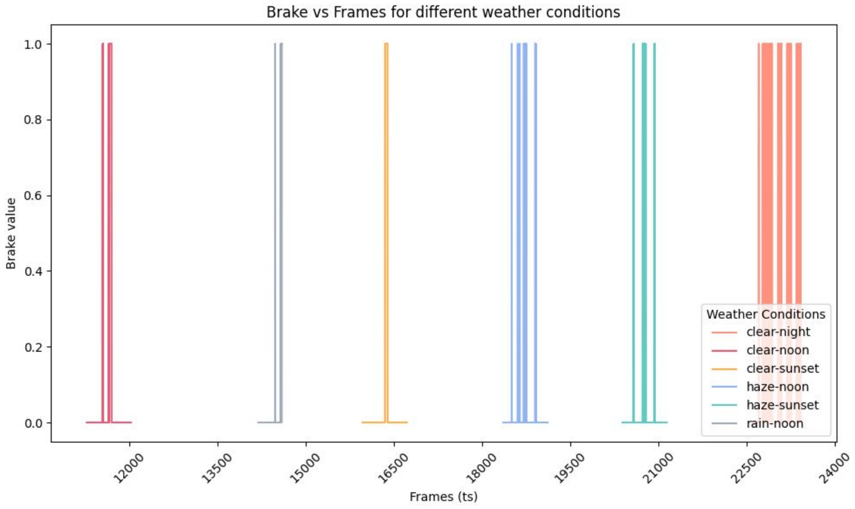
- throttle



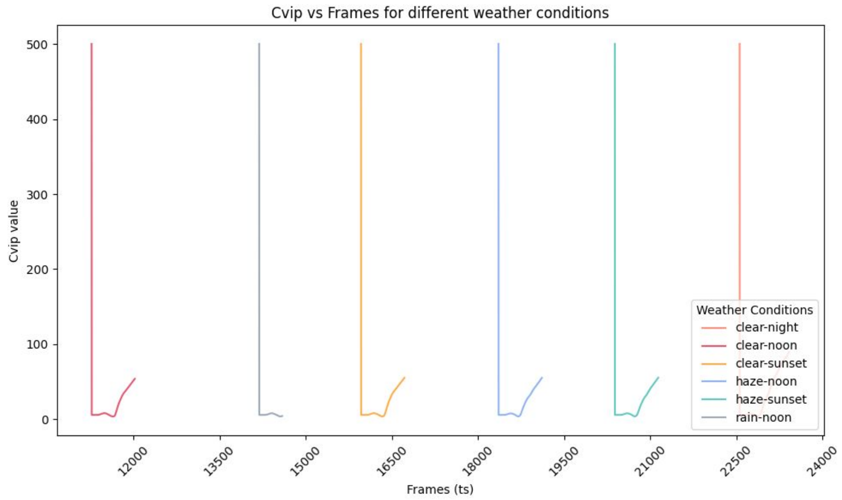
- steer



- brake



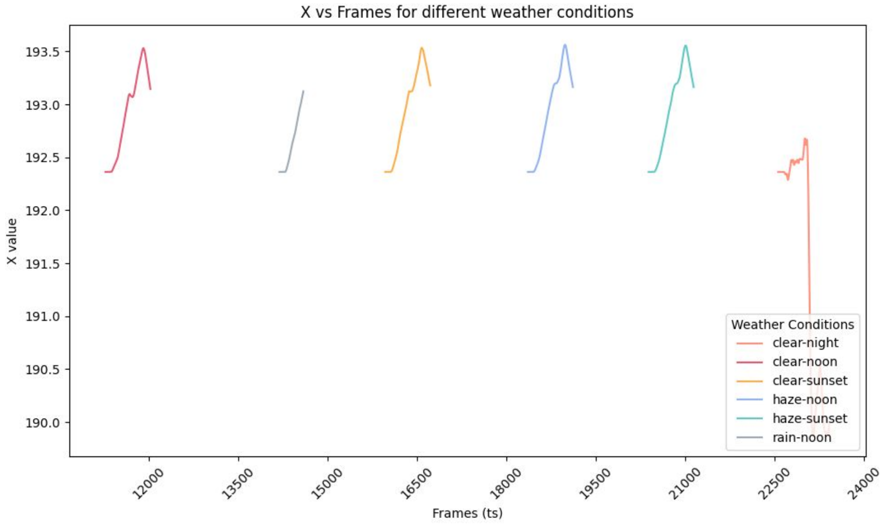
- cvip



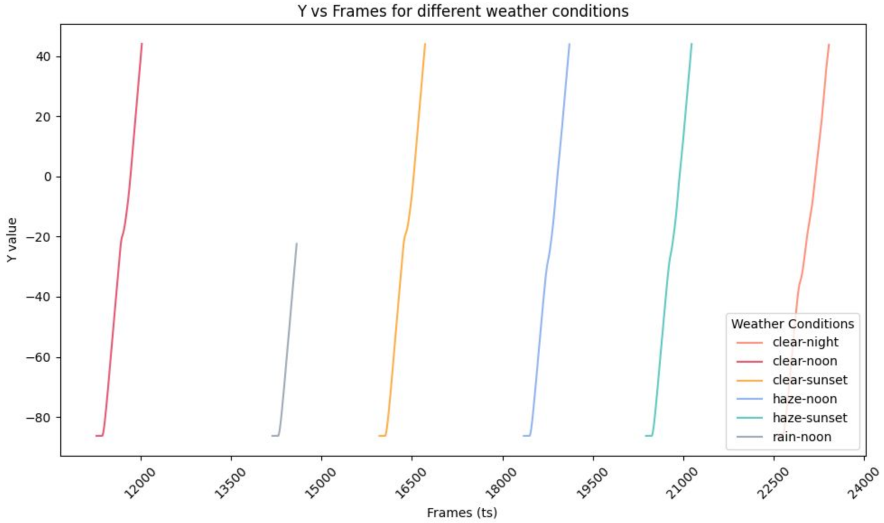
Task 1

Q3: Campaign result visualization

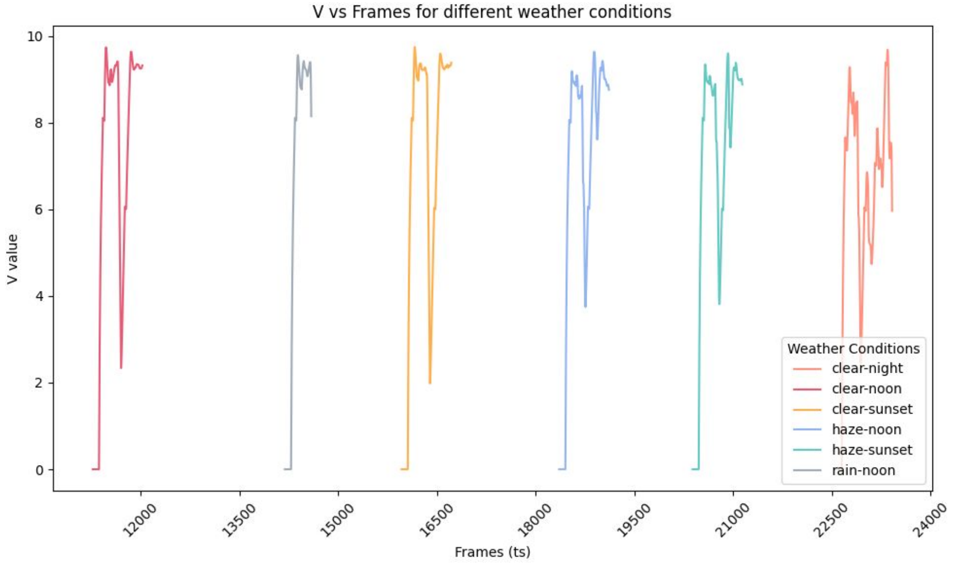
- x



- y



- v



Task 1

Q4: Based on your intuition and life experience, which of the features do you think will change during an accident? How will the feature(s) change?

1. **Throttle**: The throttle may decrease sharply just before the accident, as drivers usually slow down or stop when sensing a collision is imminent.
2. **Brake**: The brake input may increase dramatically during an accident, indicating the driver’s attempt to stop the vehicle.
3. **Steering**: The steering may show rapid changes during an accident, as the driver might try to swerve to avoid an object or another vehicle.
4. **Speed(v)**: Speed will typically decrease and may even reach zero as the vehicle comes to a stop due to the accident.
5. **Distance to other objects (cvip)**: The distance between the vehicle and other vehicles or objects will decrease rapidly as the accident approaches.
6. **x (horizontal direction)**: If the driver or the autonomous system attempts to swerve before the accident, x may undergo rapid changes.
7. **y (driving direction)**: If the vehicle decelerates or stops before the accident, y will gradually decrease and may eventually approach zero.

Task 1

Q4: By looking at the plots you generated in Task 1.3, combined with your reasoning (without looking at ‘route_highway.txt’), which weather condition(s) has an accident?

Based on the observations:

1. The **brake** value (brake = 1) remains at 1, indicating the vehicle applied full braking.
2. The **speed (v)** drops and does not increase again, suggesting the vehicle may have come to a stop.
3. The **cvip** value equals 0, meaning the distance between the vehicle and another object or vehicle is zero, implying a collision.

From these points, it can be inferred that an accident occurred under the "**rain-noon**" weather condition.

Task 2

Task 2

Q1: Suppose each simulation run has a result of accident/non-accident, calculate the probability of accident (counts, marginal probability)

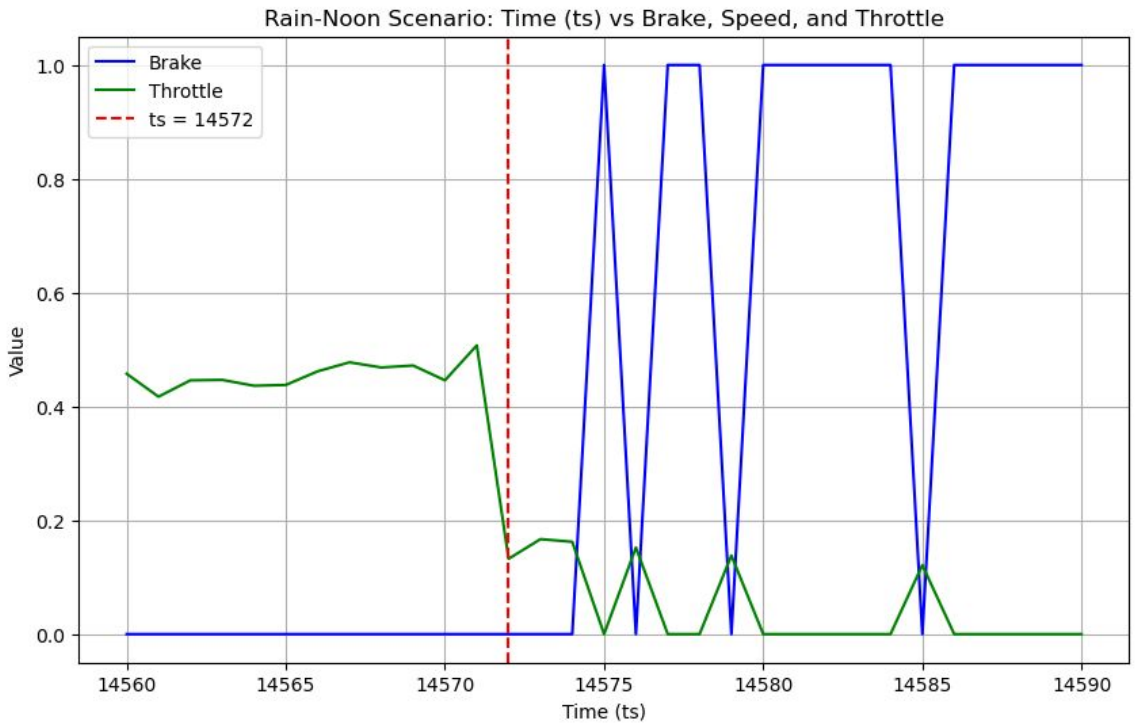
Ans:
Total number of distinct simulation runs: 6
Number of accident runs: 1
Accident_probability = accident_runs / total_runs

=> Accident probability: 0.1667

Task 2

Q2: By looking at the completion records and the plots you generated in Task 1(Discuss each accident case separately.)

- Under which weather condition(s) did the accident happen?
  - Accident occurred under the **rain-noon** condition.
- Does that match your guess in Task 1?
  - Yes, it's does.
- When did accident happen during those simulation runs?
  - We believe the accident happened at **ts = 14572**
- Why do you think the accident happened at that instance?
  - At ts = 14572, there was a sudden drop in throttle, followed by a pattern of rapid braking and releasing, which might indicate that the car's ABS (Anti-lock Braking System) was activated. Eventually, the vehicle came to a complete stop at ts = 14591.

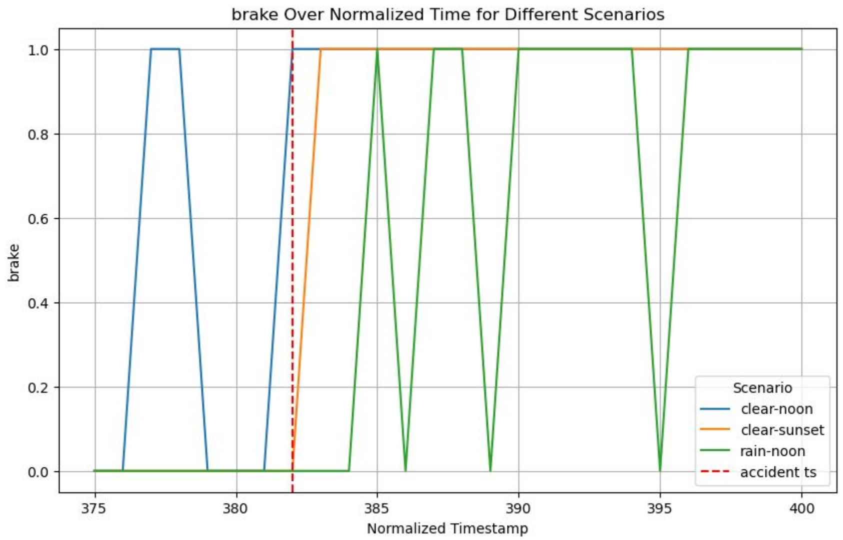


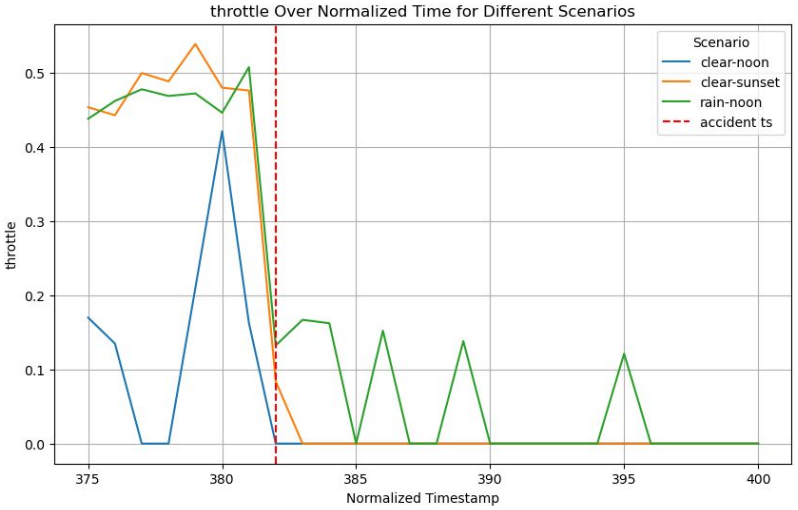
Task 2

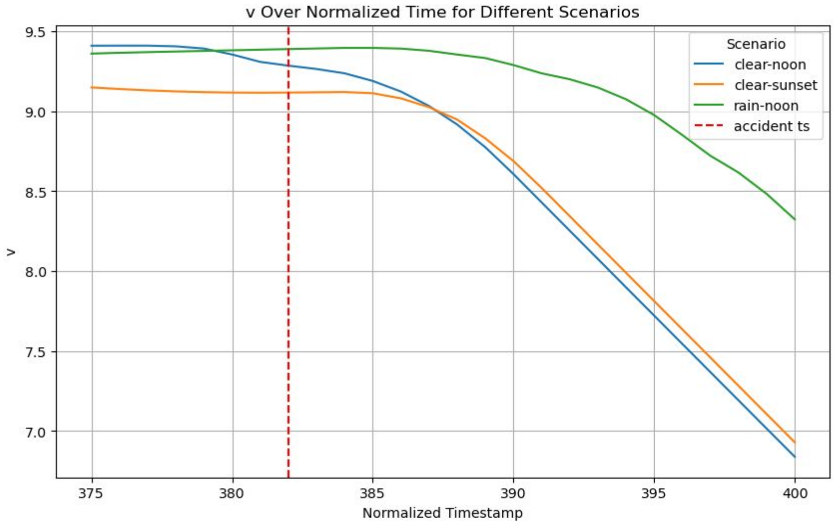
Q3: From the plots you generated in Task 1.3, do you observe any other abnormal behavior? If so, what do you think is (are) the cause(s) of this behavior?

To analyze potential abnormal behaviors, we examined the periods immediately before and after the accidents. We focused on the three scenarios: rain-noon, clear-noon, and clear-sunset, as they share similar x and y axes, making them suitable for comparison (see task 2 Q3 appendix).

In the rain-noon scenario, we observed that the AV applied the brakes later than in the other two scenarios. Furthermore, the braking pattern in rain-noon was intermittent, and the vehicle continued to accelerate, suggesting that the system may not have been properly calibrated, which prevented the vehicle from reducing its speed effectively. This can be seen in the speed (v) plot, where the rain-noon scenario did not decelerate as expected compared to clear-noon and clear-sunset.



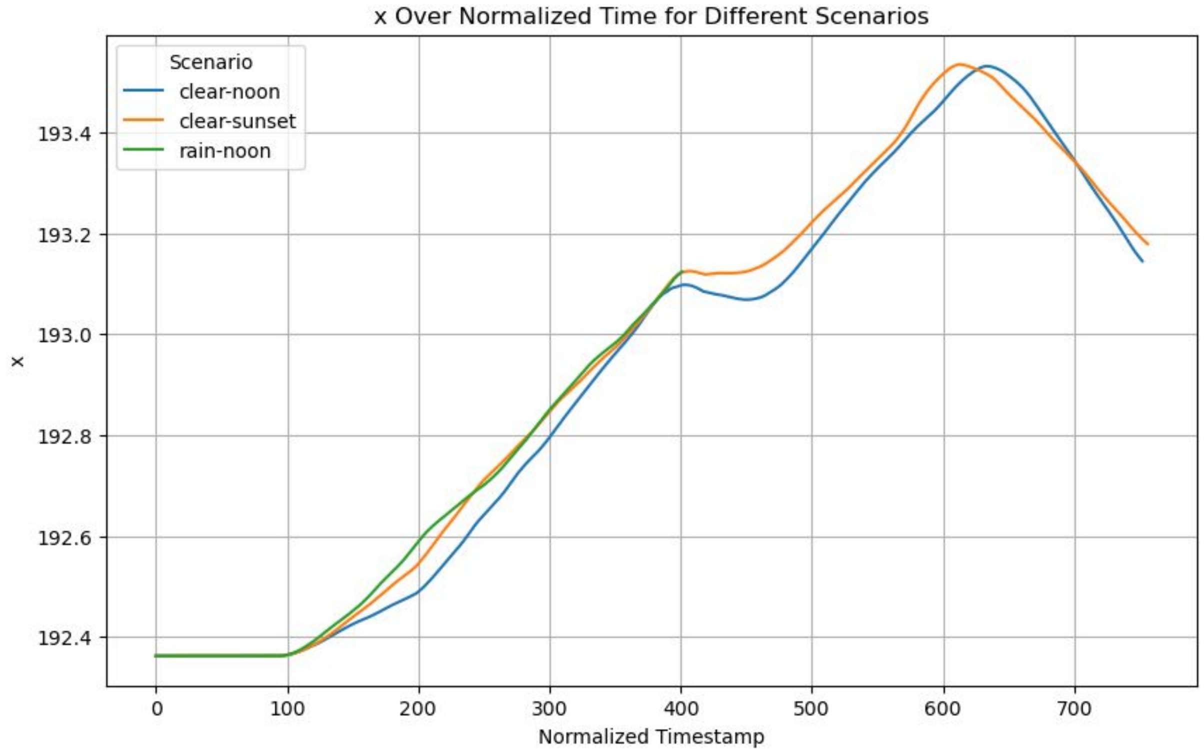


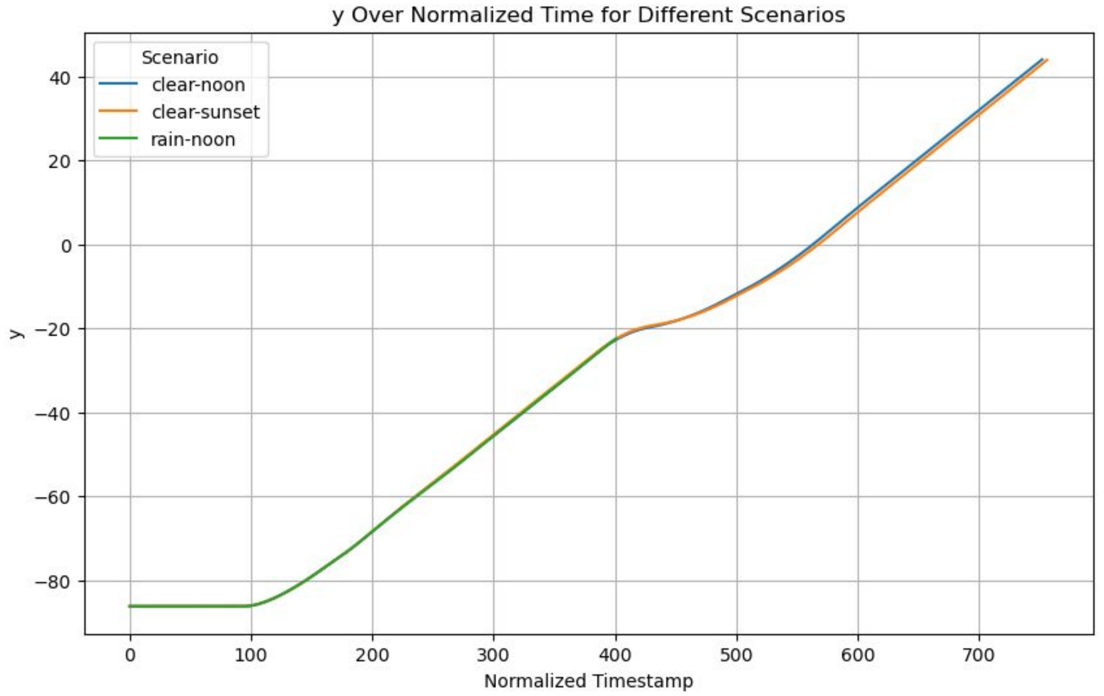


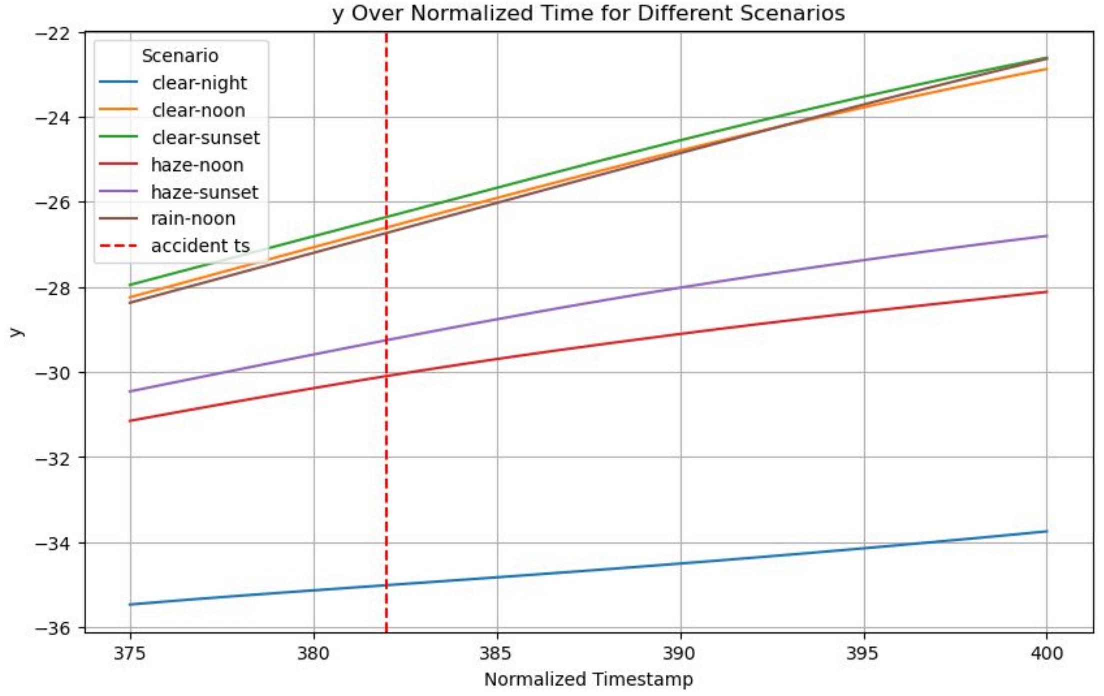
* Note that we have normalized the timestamps, meaning the starting point for each scenario is set to 0.

Task 2

Q3: Appendix



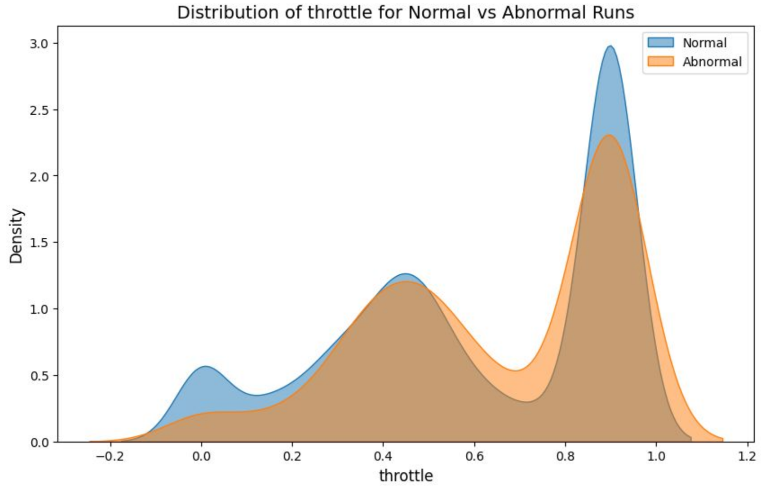




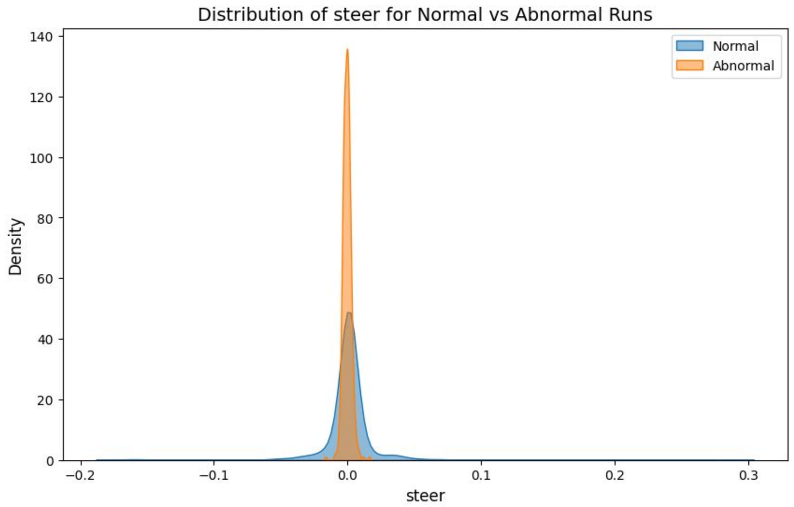
Task 2

Q4.a: Distribution of the features: abnormal (including accidents) vs normal

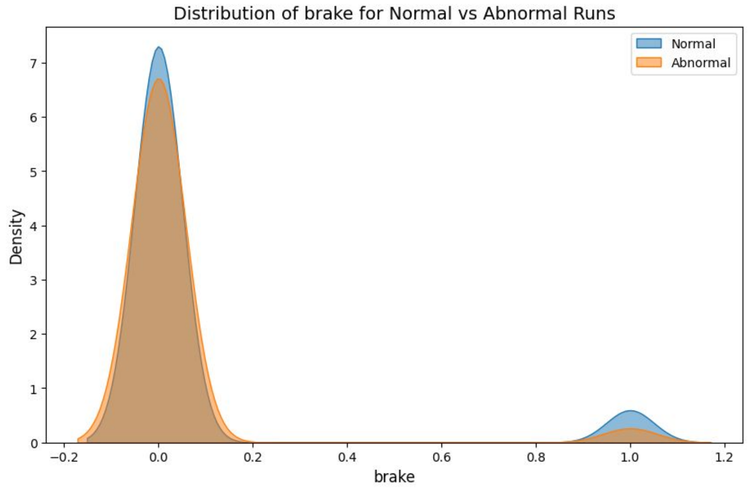
- throttle



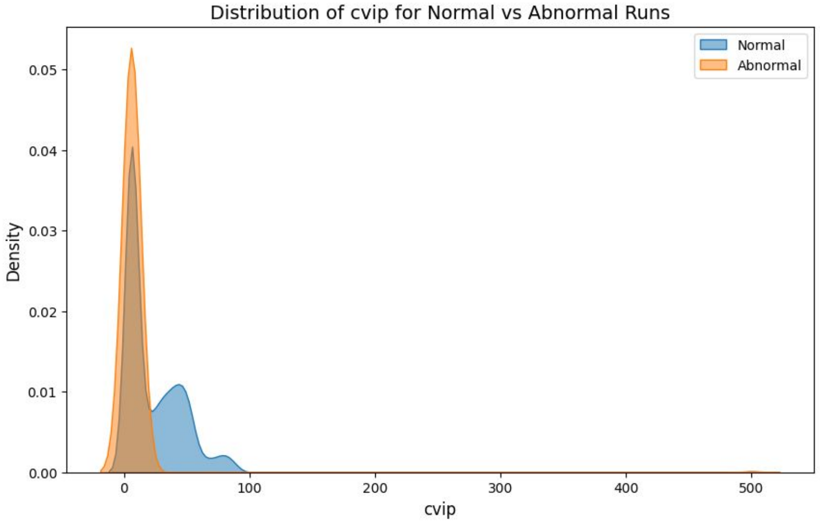
- steer



- brake



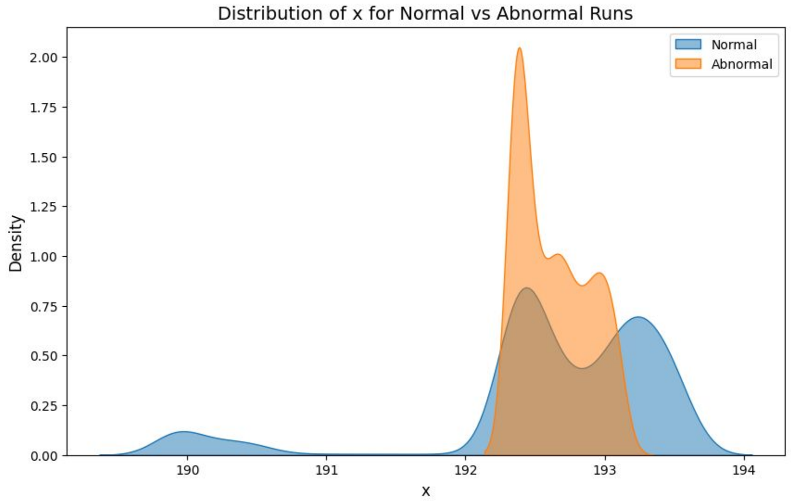
- cvip



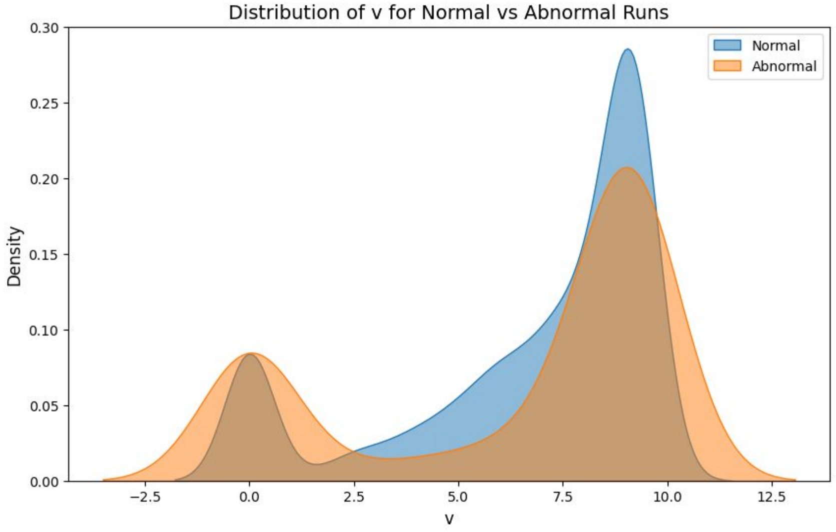
Task 2

Q4.a: Distribution of the features: abnormal (including accidents) vs normal

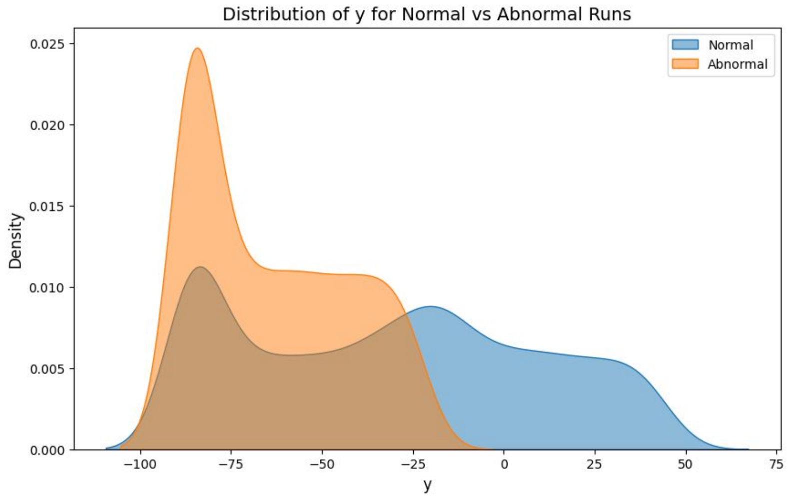
- x



- v



- y



Task 2

Q4.b: Use 2-sample t-test to test on the **'steer'** values of **abnormal runs vs normal runs.**

i. State the null and alternative hypotheses

   H0: μ abnormal = μ normal

   H1: μ abnormal ≠ μ normal

ii. Perform the test and calculate test statistics.

   We use Levene's test to check whether the variances between two groups of data (normal and abnormal steer values) are equal. Checking the equality of variances is important before performing a t-test, as it determines which version of the t-test should be used. It is a robust test for equality of variances and works well even when the data is not normally distributed.

   Levene's test result: Based on the result (Statistic: 38.9101, p-value: 0.0000), we chose Welch's t-test for further analysis.

   T-Test (Welch's T-test) result: T-test statistic: 3.9279, p-value: 0.0001

iii. Assume a significance level of 0.05, what is your conclusion?

   Based on the results of both the Levene's test and Welch's t-test, we reject the null hypothesis (H_0). Therefore, we conclude that the steering behavior (steer values) during abnormal runs is significantly different from that during normal runs. This difference is statistically significant at the 0.05 significance level.

Task 2

Q4.c: Does the testing result contradict your observation on the ”steer” feature in part 4.a? Why?
Observation in 4.a

- In 4.a, we observed the density distribution of the "steer" feature for both normal and abnormal runs:
    - For abnormal runs, the "steer" values had a narrow distribution with a mean close to zero and a small standard deviation (0.0031). This indicates that the autonomous vehicle made fewer steering adjustments in abnormal situations.
    - For normal runs, the "steer" values had a wider distribution, a slightly higher mean, and a larger standard deviation (0.0284). This suggests that the vehicle made more steering adjustments during normal operation, likely adapting to various driving scenarios.

Comparison with t-Test Result

- The t-test result showed a significant difference in the mean "steer" values between normal and abnormal runs, which is consistent with the observations made in 4.a. Both the visual inspection and the statistical test point to the same conclusion: during abnormal runs, the vehicle's steering behavior is more conservative and stable compared to normal runs.

Summary

- The observations in 4.a showed that the steering behavior is different between normal and abnormal runs, with abnormal runs displaying less variation in steering. The t-test results in 4.b support this observation by statistically confirming the significant difference in means. Therefore, the t-test results do not contradict the observations; rather, they reinforce the conclusion that the vehicle's steering behavior is significantly different between normal and abnormal conditions.

Task 2

Q5: Some of the features are better indicators of abnormal AV behavior, can you identify them?

a. By looking at the distribution plots of the features in Task 2.4, explain your choice of indicators.

Steer:

- As observed in the previous analysis, the "steer" feature showed a distinct difference in distribution between normal and abnormal runs. In abnormal runs, the "steer" values were more concentrated around zero with a lower standard deviation. This indicates that the AV made fewer steering adjustments during abnormal situations, which could signify a reduced response to changing road conditions or obstacles, making it a strong indicator of potential issues.

v (Speed):

- The distribution of "v" (velocity or speed) also showed notable differences between normal and abnormal runs. During abnormal runs, the speed was generally lower or more inconsistent compared to normal runs. This reduction or fluctuation in speed could indicate the AV’s hesitation or failure to maintain proper speed in challenging situations, making it a good indicator of abnormal behavior.

cvip (Distance to Risky Actor):

- The "cvip" feature, representing the distance to a risky NPC actor, showed significant variation in abnormal runs. In abnormal situations, the AV might fail to maintain a safe distance from other vehicles or pedestrians, resulting in a smaller "cvip" value. This suggests that the AV is either too close to potential hazards or unable to navigate away effectively, making it a critical indicator of abnormal behavior.

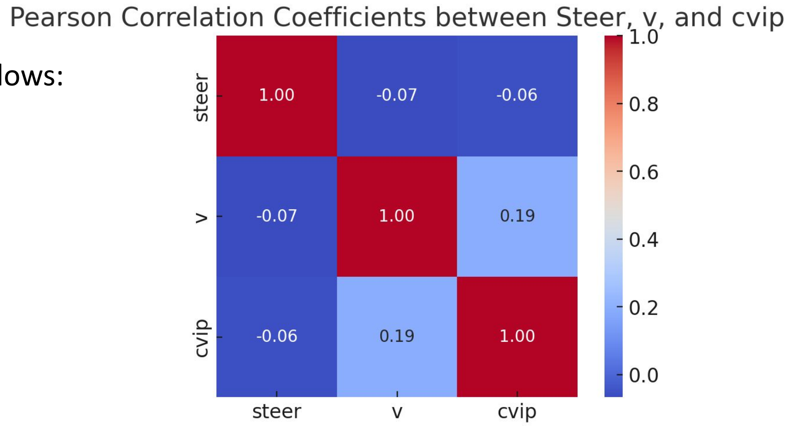
Task 2

Q5 b. For the fields you identified as good accident indicators above, are they related (Calculate the Pearson correlation coefficient between each pair of the indicators to justify your answer)? If so, how does that affect the predicting power of using one indicator versus using all of them?

Pearson Correlation Results

The Pearson correlation coefficients between the identified features ("steer", "v", and "cvip") are as follows:

- **Steer and v:** -0.067 (weak negative correlation)
- **Steer and cvip:** -0.056 (weak negative correlation)
- **v and cvip:** 0.192 (weak positive correlation)



Interpretation

1. **Low Correlation Between Indicators:**
   - The correlation coefficients between these features are relatively low, indicating that they are not strongly related to each other. This suggests that each feature captures different aspects of abnormal AV behavior independently.
2. **Implication for Predictive Power:**
   - Since the features are weakly correlated, using all of them together can provide a more comprehensive view of the AV's behavior in different abnormal scenarios. For example, "steer" might indicate issues with directional control, "v" can signal problems with speed regulation, and "cvip" reflects potential failures in maintaining safe distances.
   - Using a combination of these indicators can improve the predictive power of the model as they complement each other, covering multiple dimensions of AV behavior.
3. **Using a Single Indicator vs. All Indicators:**
   - If we rely on a single indicator, such as "steer," we might miss other critical aspects like speed anomalies or unsafe distances. Using all identified indicators provides a more robust assessment, reducing the risk of overlooking abnormal behavior that could lead to accidents.

In conclusion, while "steer," "v," and "cvip" are weakly correlated, each offers unique predictive insights. Therefore, using them collectively will enhance the overall accuracy of predicting abnormal AV behavior

Task 2

Q6: Suppose we want to use hypothesis testing to test whether the field you choose from Task2.5 is indeed a good indicator of abnormal AV behavior, using the Kolmogorov–Smirnov two-sample test.

a. Construct the null and the alternative hypothesis and state them below

H0: The distribution of the chosen feature (steer, v, or cvip) is the same for both normal and abnormal runs

H1:The distribution of the chosen feature (steer, v, or cvip) is different for normal and abnormal runs

b. Perform the KS two-sample test and calculate its statistics.

1. Steer

- **KS Statistic:** 0.2516
- **p-Value:** 8.7e-21

2. v (Speed)

- **KS Statistic:** 0.1587
- **p-Value:** 1.8e-08

3. cvip (Distance to Risky Actor)

- **KS Statistic:** 0.4971
- **p-Value:** 3.1e-83

c. Assume a significance level of 0.05, what is your conclusion?

For each of the features (steer, v, cvip), the p-values are significantly smaller than the significance level of 0.05. This means we reject the null hypothesis (H0) for all three features. In other words, there is a significant difference in the distribution of each feature between normal and abnormal runs.

This result confirms that "steer", "v", and "cvip" are indeed good indicators of abnormal AV behavior as their distributions differ significantly between normal and abnormal scenarios.

Task 2

Q6:

d. Repeat the same test on a feature that you did not select as an indicator of abnormal behavior in Task 2.5, what is your conclusion?

1. **Brake**
   - **KS Statistic:** 0.0379, **p-Value:** 0.6568
   - **Conclusion:** No significant difference in distribution between normal and abnormal runs. "Brake" is not a good indicator of abnormal behavior.
2. **y (Vertical Position)**
   - **KS Statistic:** 0.4404, **p-Value:** 1.4e-64
   - **Conclusion:** Significant difference in distribution, indicating "y" is a strong indicator of abnormal behavior.
3. **x (Horizontal Position)**
   - **KS Statistic:** 0.3340, **p-Value:** 1.2e-36
   - **Conclusion:** Significant difference in distribution, making "x" a good indicator of abnormal behavior.
4. **Throttle**
   - **KS Statistic:** 0.0892, **p-Value:** 0.0057
   - **Conclusion:** Significant difference in distribution. "Throttle" can be a useful indicator, though less pronounced.

Summary

Among the non-selected features, "brake" does not show significant differences in distribution and is not a good indicator. However, "x", "y", and "throttle" all exhibit significant differences between normal and abnormal runs, indicating they may also serve as useful indicators of abnormal AV behavior

Task 2

Q6:

e. What are the major differences between the KS test and the t-test?

1. **Type of Test:**
   - The **t-test** is a parametric test that compares the means of two groups to determine if they are statistically different. It assumes that the data follows a normal distribution and that the variances are equal (for the standard t-test).
   - The **KS test** is a non-parametric test that compares the distributions of two groups. It does not make any assumptions about the distribution of the data and is more versatile in comparing differences across the entire distribution, not just the mean.
2. **Purpose:**
   - The **t-test** specifically tests whether there is a significant difference in the mean values of a feature between two groups.
   - The **KS test** tests whether there is a significant difference between the cumulative distribution functions (CDFs) of two groups, identifying differences in shape, spread, or location of the distributions.
3. **Sensitivity:**
   - The **t-test** is sensitive to changes in the mean but might not detect differences in distribution shapes, tails, or outliers.
   - The **KS test** is sensitive to changes in the overall distribution, including differences in medians, variances, and distribution shapes.
4. **Applicability:**
   - The **t-test** is suitable when comparing means under the assumption of normality and similar variances.
   - The **KS test** is suitable for comparing any two distributions and is more robust in cases where data do not meet the assumptions of the t-test.

In summary, the KS test provides a broader comparison of the distributions between two groups, while the t-test focuses solely on differences in the means.

Task 2

Q7: Keeping in mind that this experiment is executed over a period of time, what assumption did you make when using the KS two-sample test on the distributions in Task 2.6? Are you able to come up with one situation where this assumption fails?

When applying the KS two-sample test in Task 2.6, the primary assumption made is that the observations for both normal and abnormal runs are independent and identically distributed (i.i.d.) samples from their respective distributions. This means:

1. **Independence**: The value of the feature at each time point is assumed to be independent of the values at other time points.
2. **Stationarity**: The distributions of the features are assumed to remain consistent over time, without significant shifts or trends that could affect their behavior.

Potential Situation Where This Assumption Fails

Scenario: Temporal Correlation or Trends Over Time

If the AV behavior is affected by temporal factors such as changing weather conditions, road congestion, or sensor fatigue over time, the independence and stationarity assumptions may not hold. For example:

- **Example Situation:** Suppose during a long test drive, the AV's sensors gradually degrade due to dirt accumulation or wear, affecting the vehicle's response time and steering precision over the period of the test. This degradation introduces a time-dependent trend, violating the assumption of independent and identically distributed observations. As a result, the KS test may incorrectly attribute these systematic temporal changes to abnormalities in AV behavior rather than recognizing them as a gradual degradation over time.

In such cases, using a KS test without accounting for these time-dependent factors could lead to misleading conclusions about the AV's performance.

Task 2

Q8: The dynamic-time-wrapper (DTW) is a method to compare two time-series data (such as the control and the trajectory data collected in our simulation). Use the DTW package in python (dtaidistance · PyPI), and apply the DTW distance on the two time-series dataset (using steering data of clear-noon as a reference): (1) steering data of clear-night and (2) steering data of clear-sunset. What can you say about the DTW distance for (1) and (2) with respect to the reference?

1. **DTW Distance Analysis:**
   - **Clear-Noon vs Clear-Night: 1.512**
     - The larger DTW distance of 1.512 suggests that the steering behavior in the clear-night scenario is quite different from that of the clear-noon scenario. This indicates that the AV's steering response varies significantly between daytime and nighttime conditions, potentially due to factors like reduced visibility or sensor performance in low-light conditions.
   - **Clear-Noon vs Clear-Sunset: 0.059**
     - The much smaller DTW distance of 0.059 indicates that the steering behavior in the clear-sunset scenario closely resembles the clear-noon scenario. This suggests that the AV's steering performance is more consistent between noon and sunset, possibly because both scenarios offer similar lighting conditions, making it easier for the AV to maintain stable behavior.
2. **Comparative Analysis:**
   - **Steering Stability:** The clear-sunset scenario shows a much closer alignment with the clear-noon scenario compared to the clear-night scenario. This suggests that the AV operates more reliably and with consistent steering patterns in well-lit conditions, while its behavior may become more unpredictable or cautious at night.
   - **Implication:** These results highlight the impact of lighting conditions on AV steering behavior. The significant difference between clear-noon and clear-night steering patterns may necessitate additional considerations for AV performance and safety under low-light conditions.

Overall, the DTW distance results confirm that the steering behavior of the AV is more consistent between noon and sunset, while it changes noticeably under nighttime conditions.