Data Science applied to maintenance planning optimization

Summary

Situation

A new data science consulting company was hired to solve and improve the maintenance planning of an outsourced transport company. The company maintains an average number of trucks in its fleet to deliver across the country, but in the last 3 years it has been noticing a large increase in the expenses related to the maintenance of the air system of its vehicles, even though it has been keeping the size of its fleet relatively constant. The maintenance cost of this specific system is shown below in dollars:
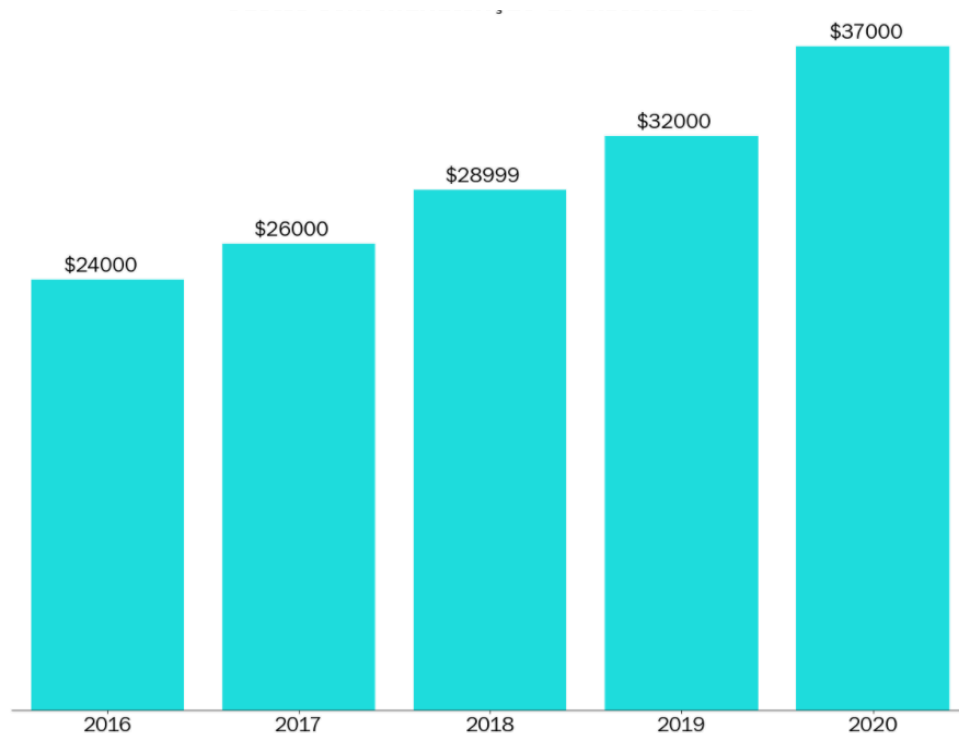
Your objective as a consultant is to decrease the maintenance costs of this particular system. Maintenance costs for the air system may vary depending on the actual condition of the truck.

- If a truck is sent for maintenance, but it does not show any defect in this system, around $10 will be charged for the time spent during the inspection by the specialized team.

- If a truck is sent for maintenance and it is defective in this system, $25 will be charged to perform the preventive repair service.

- If a truck with defects in the air system is not sent directly for maintenance, the company pays $500 to carry out corrective maintenance of the same, considering the labor, replacement of parts and other possible inconveniences (truck broke down in the middle of the track for example).

During the alignment meeting with those responsible for the project and the company's IT team, some information was given to you:

- The technical team informed you that all information regarding the air system of the paths will be made available to you, but for bureaucratic reasons regarding company contracts, all columns had to be encoded.

- The technical team also informed you that given the company's recent digitization, some information may be missing from the database sent to you.

Finally, the technical team informed you that the source of information comes from the company's maintenance sector, where they created a column in the database called **class**: "pos" would be those trucks that had defects in the air system and "neg" would be those trucks that had a defect in any system other than the air system.

Those responsible for the project are very excited about the initiative and, when asking for a technical proof of concept, they have put forth as main requirements:
- Can we reduce our expenses with this type of maintenance using AI techniques?

- Can you present to me the main factors that point to a possible failure in this system?

These points, according to them, are important to convince the executive board to embrace the cause and apply it to other maintenance systems during the year 2022.

About the database

Two files will be sent to you:
- *air_system_previous_years.csv*: File containing all information from the maintenance sector for years prior to 2022 with 178 columns.
- *air_system_present_year.csv*: File containing all information from the maintenance sector in this year.
- Any missing value in the database is denoted by *na*.

The final results that will be presented to the executive board need to be evaluated against *air_system_present_year.csv*.

Challenge Activities

To solve this problem we want you to answer the following questions:

1. What steps would you take to solve this problem? Please describe as completely and clearly as possible all the steps that you see as essential for solving the problem.

2. Which **technical** data science metric would you use to solve this challenge? Ex: absolute error, rmse, etc.

3. Which business metric **would** you use to solve the challenge?

4. How do technical metrics relate to the business metrics?

5. What types of analyzes would you like to perform on the customer database?

6. What techniques would you use to reduce the dimensionality of the problem?

7. What techniques would you use to select variables for your predictive model?

8. What predictive models would you use or test for this problem? Please indicate at least 3.

9. How would you rate which of the trained models is the best?

10. How would you explain the result of your model? Is it possible to know which variables are most important?

11. How would you assess the financial impact of the proposed model?

12. What techniques would you use to perform the hyperparameter optimization of the chosen model?

13. What risks or precautions would you present to the customer before putting this model into production?

14. If your predictive model is approved, how would you put it into production?

15. If the model is in production, how would you monitor it?

16. If the model is in production, how would you know when to retrain it?