POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

SURVIVAL ANALYSIS TO SUPPORT MAINTENANCE FOR AEROSPACE AGENCIES

Project for Nonparametric Statistics Course - Mathematical Engineering

Simone Ciardulli, Celeste De Bernardinis, Darvin Koka, Federica Rena

Academic year:
2023-2024

Abstract: This report introduces a new approach that applies Survival Analysis in order to provide a more precise prediction of satellite lifetimes to space agencies. Aerospace industries have long relied on outdated models to predict the lifespan of their satellites, leading to significant uncertainties in maintenance schedules and potential loss of satellites, resulting in substantial financial implications. Our startup has developed a new solution that enables preventive maintenance, thereby mitigating the risk of premature satellite loss. Furthermore we demonstrate why the American market is the most suitable for starting our business.

Key-words: Survival Analysis, Satellites, Predicted lifetime, Preventive Maintenance

1. Introduction

The satellite market is growing and profitable, making it necessary to have an accurate model that estimates satellite lifetimes and establishes preventive maintenance interventions to avoid the loss of a satellite.

We have discovered through appropriate tests that the effective lifetime of many retired satellites has been very different from the one expected by space agencies at the moment of the launch. This discrepancy suggests that in the current landscape, the models used to predict satellite lifespans are significantly imprecise. This lack of precision poses a risk for space agencies, as they may lose their satellites due to an inability to accurately schedule maintenance.
Specifically, the lack of accurate predictive models results in uncertainty about when a satellite will fail. The risk that this poses for space agencies is not being able to retrieve the satellite in time for repairs. If it's too late, the only option left for space agencies is to send their satellites in a so-called "Graveyard Orbit", resulting in permanent loss.

However, a more precise predictive model could make space agencies more aware of the time frame of a satellite's critical failure. This would enable them to schedule preventive maintenance or control checks, thereby avoiding the loss of their satellites and saving significant costs associated with such loss.
For this reason, the aerospace sector presents a flourishing market for data scientists who can propose precise predictive models.

1.1. Stakeholders

To address this issue, we have established a consulting startup with the objective of developing a more precise predictive model. Space agencies that engage our consulting services will be equipped with the ability to accurately forecast the lifespan of their satellites. This will enable them to schedule preventive maintenance and avoid the substantial costs associated with satellite loss.

A more comprehensive examination of stakeholders can be facilitated through the use of a power-interest matrix, functioning as a visual tool for evaluating stakeholders and their individual levels of power and interest. The matrix is segmented into four quadrants, each categorizing stakeholders based on their degree of power and interest. In the context of our project, we can categorize stakeholders in the power-interest matrix as follows:

- **Space Agencies (High Power, High Interest)**
  Space agencies are key stakeholders with significant influence over and vested interest in our project. They are the primary beneficiaries of the predictive model, and their satisfaction is paramount to the success of our project;

- **Public Governments (High Power, Low Interest)**
  Public governments wield regulatory power over space activities but may have less direct interest in the specific outcomes of our project. It's crucial to keep them informed and ensure compliance with all relevant regulations. Furthermore, governments are the starting point for allocating funds, therefore it is necessary to obtain their consent as well;

- **Maintenance Service Companies (Low Power, High Interest)**
  These companies could provide the necessary maintenance services based on the predictions of our model. While they may not have significant power over the project, their high interest stems from the potential business opportunities it could generate;

- **Public Opinion (Low Power, Low Interest)**
  The general public has low power and interest in our project. However, maintaining a positive public image is always beneficial, so monitoring and managing public opinion is still important.

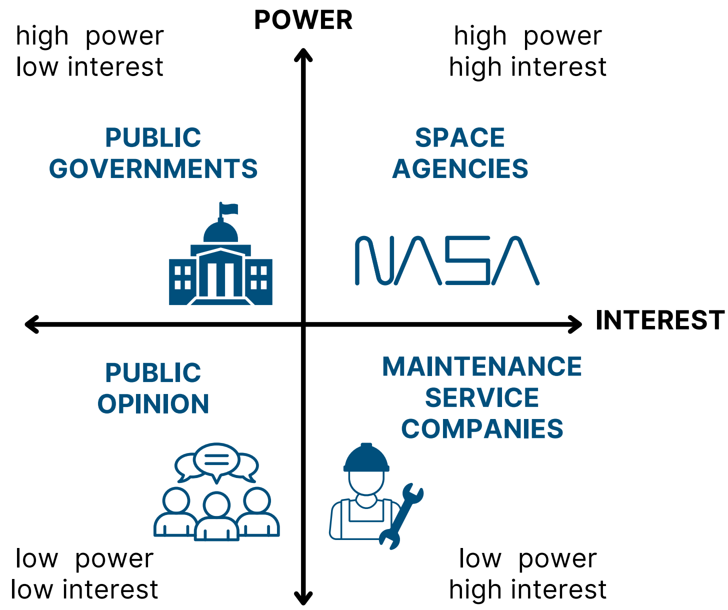


Figure 1: Power Interest matrix

Moreover, we cannot overlook the substantial costs associated with launching a startup. It's crucial for us to promote ourselves effectively and focus our resources to the most promising markets for our business. To do this, we sought to identify the markets that would benefit the most from our analysis. Therefore, we aimed to determine the continents that have, this far, produced the least accurate predictive models. This targeted approach enables us to optimize the impact of our innovative solution in the satellite market.

1.2. Research Questions

After conducting a stakeholder analysis and deciding to orient our business towards space agencies, our focus was on providing them with an effective tool to enhance preventive maintenance for satellites.
To address this objective, we sought answers to the following research questions:

- Is it feasible to formulate a more accurate model for predicting satellite lifetimes?
- When is it appropriate to initiate maintenance processes for space agencies?

For this purpose, we employed survival analysis tools.

2. Dataset

Data come from UCS Satellite Database, which contains 6718 satellites and 13 covariates, about technical information, orbit information and mission information and from SATCAT (Satellite Catalog):

Variable	Description
Launch mass	in kilograms
Date of Launch	from 29/09/1988 to 21/10/2022
Expected Lifetime	in years
Apogee	in kilometers
Perigee	in kilometers
Period	in minutes
Inclination	in degrees
Eccentricity	measure of how circular (round) or elliptical (squashed) an orbit is
Class of Orbit	2 levels: Low Earth Orbit (LEO), Geosynchronous Orbit (GEO)
Purpose	4 levels: Communications, Earth Observation, Space Science, Technology Development
Users	4 levels: Civil, Commercial, Government, Military
Country	the country of satellite's operator

Table 1: The covariates.

2.1. Dataset Cleaning and Preprocessing

At the beginning of our analysis, we only considered the UCS database [4].
The preprocessing and data cleaning phase was the initial step of our analysis and proved to be quite time-consuming due to the structure and type of data.

- **Removal of Irrelevant Information**
  We began by removing information that was not pertinent to our objective, such as alternative names of satellites or the launch site;
- **Handling Missing Data**
  Next, we analyzed the missing data and removed features with a high number of missing values. We also eliminated satellites that lacked essential information for our analysis, such as the launch date, ending up with a total of 382 satellites;
- **Resolving Conflicts in Categorical Variables**
  Another critical issue with the dataset was that some satellites belonged to multiple levels of the same categorical variable. There were conflicts particularly for the variables "Country", "Users", and "Purpose". To eliminate ambiguity, we retained only the first level reported for each variable. For instance, for a satellite that had USA and Canada as the value of "Country", we retained only USA;
- **Annotation of Satellite Status**
  The most critical problem was that the dataset lacked the status of the satellites. We did not have a

variable that indicated whether the satellite was still active, censored, or retired, and most importantly, a "Final Date" variable that accounted for the satellite's decommissioning or censoring date. To address this deficiency, we manually annotated each remaining satellite from the initial dataset cleanup, seeking information from various official sources and performing double checking among them, primarily referring to the site Celestrak [5]. For each satellite, the "Final Date" variable was introduced with the following criterion:

  - "Final Date" = "Disposal Date" if the satellite was "retired"
  - "Final Date" = "Censoring Date" if the satellite was "censored"
  - "Final Date" = November 17, 2023, i.e., the start date of the analysis, if the satellite was "active"

  Moreover, we flagged as censored the active satellites;

- **Creation of the "Effective Lifetime" Variable**
  After this long and meticulous work, we created the "Effective Lifetime" variable, computed as the difference in years between the final date and the launch date;

- **Grouping Satellites by Continent**
  Finally, we noticed that grouping satellites by country resulted in too many levels with very few observations, such as one or two. We, therefore, decided to group satellites by continent in order to perform a meaningful analysis.

This comprehensive and rigorous approach to data preprocessing and cleaning has laid a solid foundation for our subsequent analysis, highlighting the importance of meticulous data handling in extracting meaningful insights.

2.2. Integration of an Additional Dataset

In the context of survival analysis, obtaining a sufficiently adequate and valid model necessitates a high number of events. Specifically, simulation work has suggested that at least 10 events need to be observed for each covariate considered. Anything less will lead to problems.

Unfortunately, our initial dataset, the UCS database [4], did not meet this requirement, hence the need to find additional observations. The problem was resolved by integrating the observations already present with those in the SATCAT dataset [5].

The two datasets contained similar but not identical information, for example the set of covariates was not the same. It is worth emphasizing that a non-trivial data manipulation task was carried out, following the same procedure as for the first dataset.

Only observations present in both datasets and that were incomplete in the UCS database were considered, manually searching once again the missing information needed for the analysis, i.e. the disposal dates and the status. This resulted in a dataset made of 714 observations. A set of covariates was chosen as the intersection of the two sets of covariates from the two different datasets. The considered set of variables is the one listed in the table 1.

Finally, since one of our goals was to create a model and validate it, we splitted the dataset into a training set and a test set. Obviously, the training set was used to train the model, and the test set was used to implement the tests in order to identify the differences between our model and the existing one.

3. Dataset Inspection

To underscore the unreliability of the current models for space agencies, we conducted a nonparametric paired test to assess the significant difference between the effective lifetimes and the declared expected lifetimes.

3.1. Is the Existing Model Reliable?

We exclusively considered retired satellites because censored observations cannot be taken into consideration, as we don't have access to the real time event for those satellites.

We used the following notation:

- n_{ret} denotes the number of retired satellites;
- y_{i1} denotes the effective lifetime of satellite i, $i = 1, \ldots, n_{ret}$;
- y_{i2} denotes the expected lifetime of satellite i, $i = 1, \ldots, n_{ret}$;
- Y_1 is the random variable that denotes the effective lifetime's observations;
- Y_2 is the random variable that denotes the expected lifetime's observations.

It's to be noted that the we have paired units across groups:

$$(Y_{i1}, Y_{i2}) \overset{iid}{\sim} (Y_1, Y_2),\ i = 1, \ldots, n_{ret}$$

The hypothesis test is:

$$H_0\ :\ Y_1 \overset{d}{=} Y_2\ vs\ H_1\ :\ Y_1 \overset{d}{\neq} Y_2$$

Under H_0 exchangeability is just between and within pairs (e.g., if we want to preserve likelihood, pairs cannot be splitted up).

The test statistics that we used is $T = |\bar{Y}_1 - \bar{Y}_2|$ and the resulted p-value of the test is 0.037, therefore the two models are significantly different at a confidence level of 95%.

It's noteworthy that the difference between the expected lifetime and the actual one fails to meet the assumption of Gaussianity. Specifically, the Shapiro-Wilk normality test yields a p-value of 0.0001523. This observation underscores the appropriateness of employing a nonparametric statistical framework for our case.

Moreover we verified this difference through the following DDplot, providing additional confirmation.

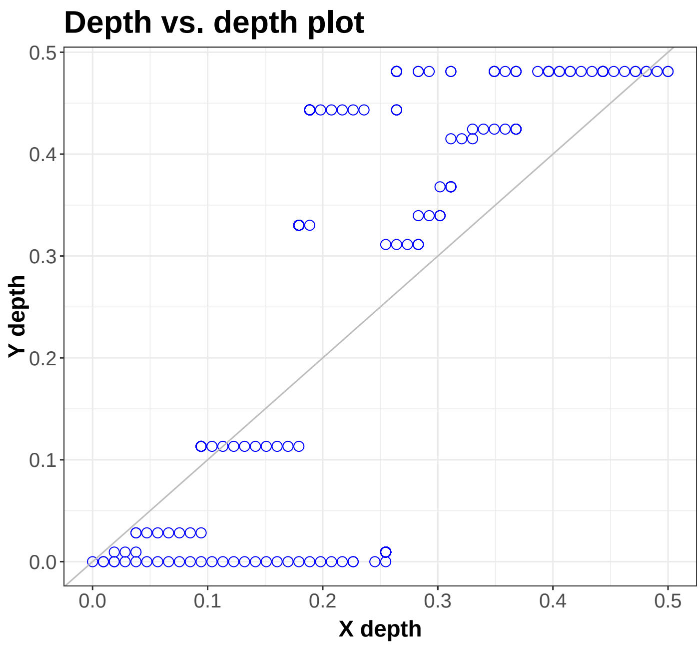


Figure 2: DDplot: effective vs expected lifetime

Additionally, we examined the disparity between the effective and the expected lifetime across different continents through the following graph:

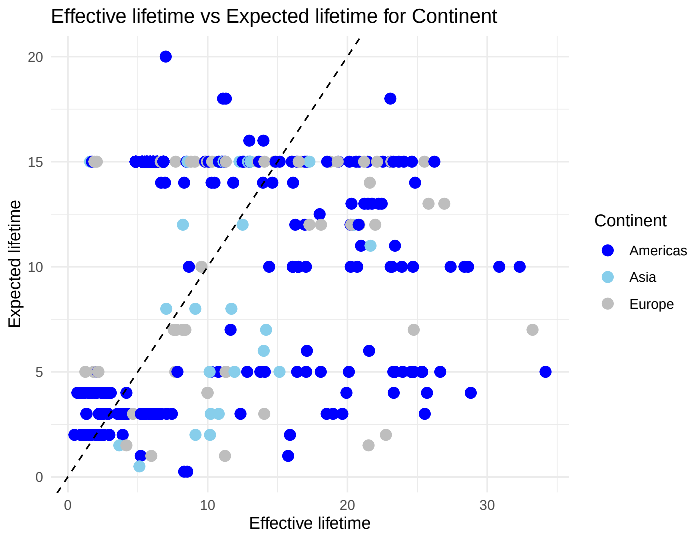


Figure 3: Effective lifetime vs Expected lifetime for continents

It was worth noting that the predicted lifetimes by Asian satellite manufacturers appear to be the most accurate when compared to the actual years of operation.

Since our main objective was to produce a model within the scope of survival analysis, we decided to dedicate some time to inspecting the dataset, particularly by analyzing the survival curves of the satellites and the differences among them based on the various groups present in our dataset.

3.2. Kaplan-Meier Estimator

We estimated the survival function using the Kaplan-Meier estimator.
The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function $S(t)$ from lifetime data.
The Kaplan-Meier survival curve is defined as the probability of surviving in a given length of time while considering time in many small intervals.
There are three assumptions used in this analysis:

1. Censoring is unrelated to the outcome. Any time satellites who are censored have the same survival prospects as those who continue to be followed.
2. The survival probabilities are the same for subjects recruited early and late in the study.
3. The events occurred at the specified times.

The Kaplan-Meier estimator of the survival function $S(t)$ is:

$$\hat{S}(t) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right)$$

where:

- j is the failure (event) index in the set $\{1, \ldots, J\}$.
- J is the total number of satellites with events.
- $0 < t_1 < \ldots < t_J < 1$ are the observed ordered times of dismission.
- p_j is the conditional probability of surviving time t_j.
- n_j is the number of active satellites just before t_j, i.e., the number of satellites at risk at time t_j.
- d_j is the number of observed events at t_j.

In our case, the survival curve is represented by the following plot:

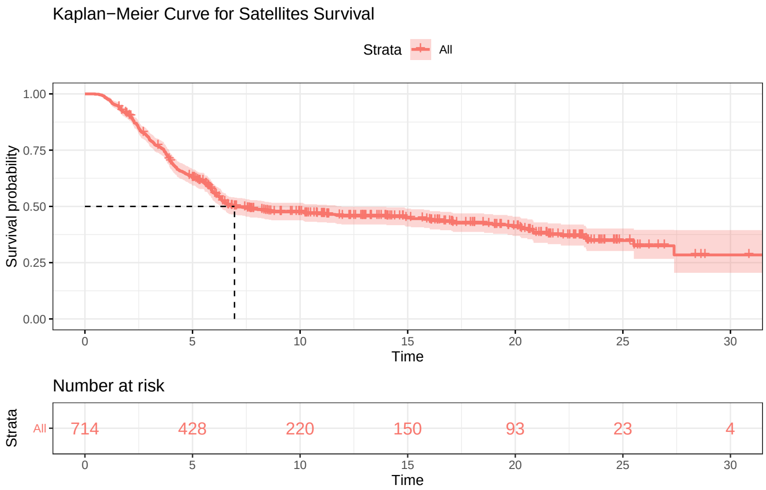
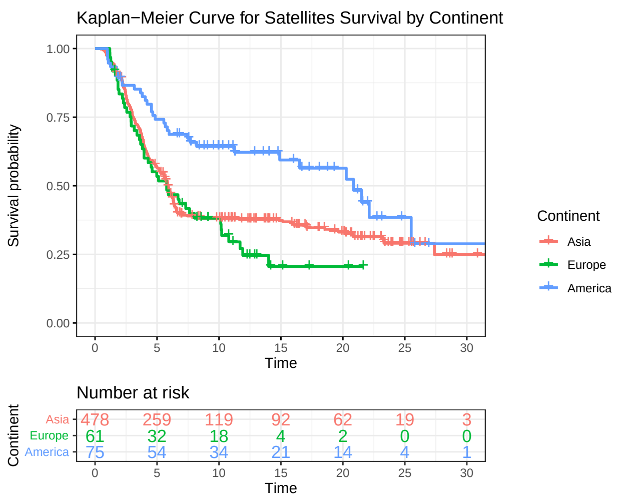
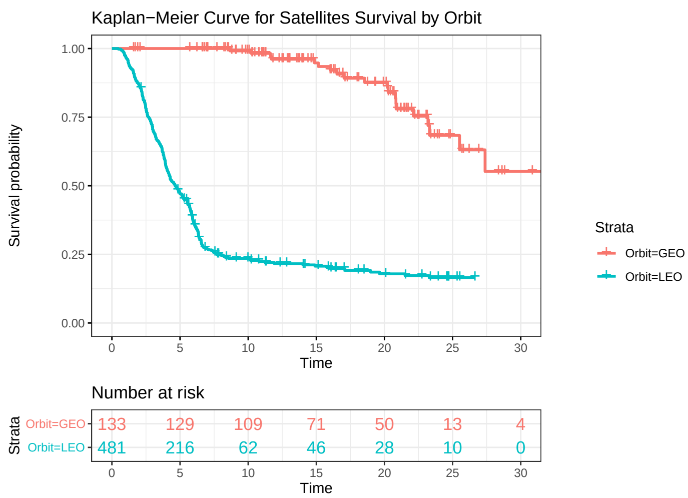


Figure 4: Survival Curve

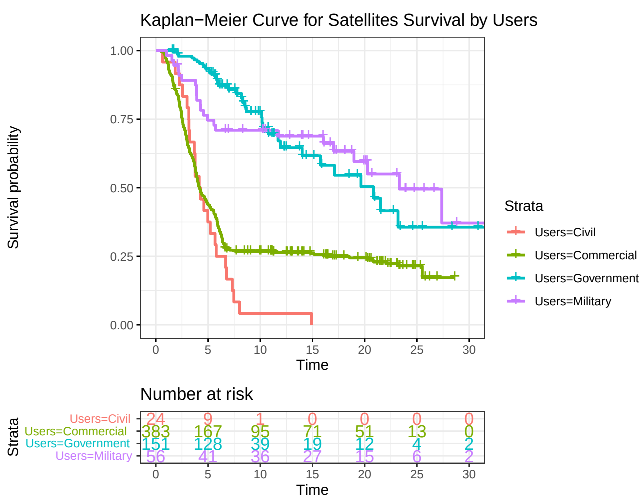
We also estimated the survival curves by Continent, Users, and Purpose. The respective curves are reported below:



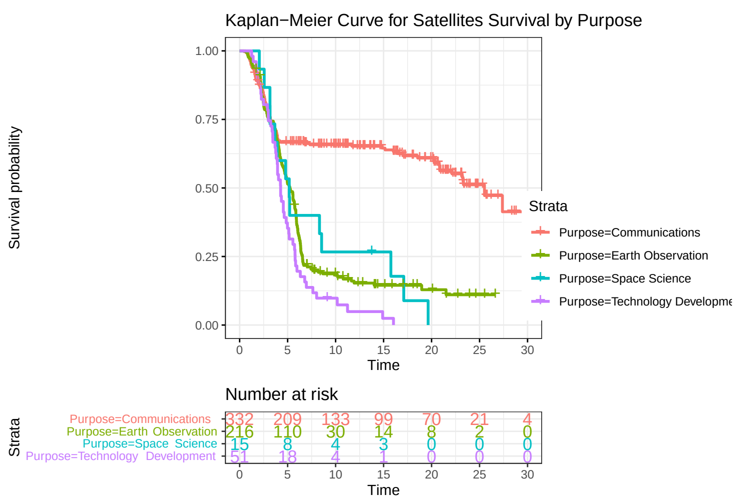
(a) Survival curves by Continent



(b) Survival Curves by Orbit



(c) Survival Curves by Users



(d) Survival Curves by Purpose

Figure 4: Survival Curves

3.3. Log-Rank Test and Hazard Ratio

Noticing differences in the survival curves among the different groups, we wanted to quantify them using the log-rank test and hazard ratio.

The Long-Rank test is:

$$H_0 : S_1(t) = \ldots = S_k(t) \quad \text{vs} \quad H_1 : \text{survival curves are not identical}$$

We obtained the following p-values:

Variable	**p-value**
Continent	0.002
Users	2.648032e-25
Purpose	1.821608e-24
Orbit	2.245295e-37

Table 2: Log-Rank p-values

So, for each group, the null hypothesis is rejected at a significance level of α =5%, and a significant difference between the survival curves is highlighted.

From the Hazard Ratios analysis Asia emerges as a risk factor compared to America and Europe, indicating that an Asian satellite is more likely to be decommissioned. Considering the satellite's use, the riskiest category is *Civil*, while the least risky is *Government*, implying that a government satellite has a lower risk of decommissioning. Concerning the satellite's purpose, the risk factors are ordered as follows from most risky to least risky: *Technology Development, Space Science, Earth Observation, Communications*. Finally, a low earth orbit (LEO) poses a higher risk than a geostationary orbit (GEO).
So, a civil Asian satellite in the field of technology development with low orbit has a very high risk of being decommissioned.

4. The model

Our primary goal in addressing the research question was to estimate the lifetimes of satellites. Specifically, by the lifetime of a satellite, we refer to the duration it remains operational, noting that a satellite may be in orbit but non-functional, categorized as orbital debris or "space junk."

The first model we implemented was a Cox model.
The Cox model is a survival analysis technique used to investigate the relationship between the time to an event (in this case, satellite failure) and predictor variables. The hazard function for the i-th satellite is given by:

$$h_i(t|\mathbf{X}_i) = h_0(t)\exp(\mathbf{X}_i^T\boldsymbol{\beta}), \quad i = 1, \ldots, n$$

Where:
- $\mathbf{X}_i \in R^p$ is the covariates vector of the i-th satellite.
- $h_0(t)$ is an unspecified non-negative function of time called the baseline hazard.
- $\boldsymbol{\beta} \in R^p$ is the vector of coefficients that we want to estimate.

The Cox Proportional Hazards (PH) model is a semiparametric model since it has the property that the baseline hazard $h_0(t)$ is an unspecified function. It is also a "robust" model since it will closely approximate the correct parametric model.

When employing a Cox model, we need to be aware that we are making the following assumptions:
- Time independence of the covariates X_i,
- Linearity in the covariates X_i,
- Additivity,
- Proportional hazards.

We initially included in our model the following variables: `Perigee`, `Apogee`, `Eccentricity`, `Inclination`, `Period`, `Mass`, `Users`, `Purpose`, `Continent`, and `Orbit`.
To ensure that our model reflected genuine explanatory variables and adhered to certain assumptions, we followed this procedure:

1. If a numerical variable proved to be statistically insignificant in explaining satellite lifetime, we removed it.
2. If a categorical variable seemed significant for most of its categories but not for all, we still chose to include the variable in the model, stratifying it based on that variable, thereby implementing a Stratified Cox PH model which is defined below.

These adjustments were made iteratively, evaluating the impact on model performance and overall significance each time a variable was added or removed.
Following this procedure, we identified the significant variables for our purpose, which are: `Apogee`, `Eccentricity`, `Mass`, `Users`, `Purpose`, `Continent`, and `Orbit`. Furthermore, the model is stratified for the Purpose and Orbit variables.

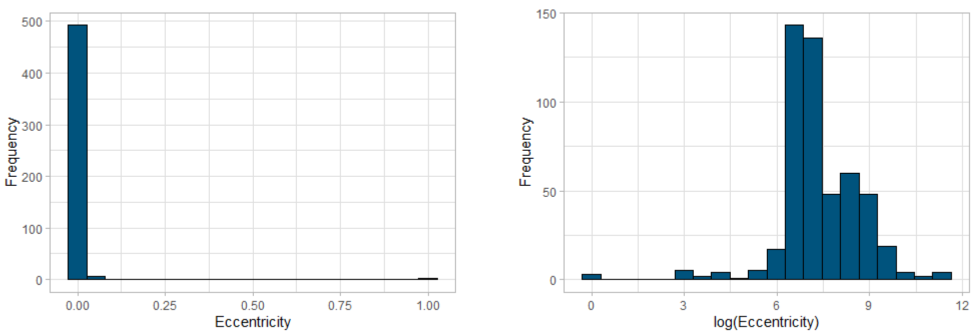
The hazard function in the Stratified Cox model is given by:

$$h_k(t|\mathbf{X}) = h_{0k}(t) \exp(\mathbf{X}^T \boldsymbol{\beta}) \tag{1}$$

Where:

- $k = 1, \ldots, K$ are the levels of the variable that is stratified.
- $h_{0k}(t)$ is an unspecified non-negative function of time called the baseline hazard for the k-th stratum.
- $\mathbf{X}$ and $\boldsymbol{\beta}$ are the vectors of covariates and coefficients, respectively.

This methodology allowed us to construct a model that accurately captures the factors influencing satellite lifetimes and provides insights into their operational durations within specific contexts. We will refer to this initial model obtained as Model 1.
Subsequently, we sought to improve this model. To do so, we scrutinized in detail the distributions of each variable found significant in the previous model. In particular, `Eccentricity` was characterized by many values all close to zero. To obtain a distribution more suitable for our purposes, we decided to consider the logarithm of `Eccentricity` in the subsequent model. Below, you can observe the difference between the two distributions:



(e) `Eccentricity` distribution.   (f) log(`Eccentricity`) distribution.

Figure 5: Distributions of the `Eccentricity`.

We then defined our final model, a Stratified Cox model, characterized by the following explanatory variables: `Apogee`, `logEccentricity`, `Mass`, `Users`, `Purpose`, `Continent`, and `Orbit`. Of course, the model is stratified with respect to `Orbit` and `Purpose`. The coefficients related to each covariate are:

Variable	coef
Apogee	-8.186e-03
log.Eccentricity	-1.727e-01
Mass	-1.385e-04
UsersCommercial	1.056e+00
UsersGovernment	-1.109e+00
UsersMilitary	8.191e-01
ContinentAsia	4.536e-01
ContinentEurope	6.647e-01

Table 3: Coefficients values

From the reported coefficients, it can be observed that having a military use and being manufactured in Europe or Asia increase the hazard of a satellite; therefore, they are risk factors.

4.1. Reverse Percentile Intervals

To further confirm the lack of significance of a specific variable, we constructed Reverse Percentile Confidence Intervals for the coefficient of the variable that we would like to remove.

A Reverse Percentile interval for θ is defined as follows:

$$2\hat{\theta} - \hat{\theta}^*_{\alpha/2} < \theta < 2\hat{\theta} - \hat{\theta}^*_{1-\alpha/2} \tag{2}$$

The quantiles of $(\hat{\theta} - \theta)$ were estimated via boostrap.

Considering a significance level $\alpha = 5\%$, we obtained the following intervals:

	lwr	pointwise	upr
Perigee	-0.0018236728	0.0001098872	0.0016278746
Inclination	-0.004559258	0.001302818	0.009182801
Period	-0.015217030	0.001295519	0.030945062

Table 4: Reverse Percentile intervals

So, it made perfect sense to remove these variables from the model since zero is included in the interval.

4.2. Assumptions and Goodness of Fit

As previously anticipated, our model must adhere to certain assumptions to be valid. Specifically, we have ensured that hazards are proportional.

We conducted the following test:

$$H_0 : \text{Hazards are proportional vs } H_1 : \text{Hazards not proportional}$$

We obtained a satisfactory result, as there is no statistical evidence to reject the null hypothesis at a significance level of $\alpha = 5\%$.

Variabile	p
Apogee	0.075
log.Eccentricity	0.675
Mass	0.887
Users	0.330
Continent	0.386
GLOBAL	0.230

Table 5: P-values of the test

In addition, to ensure that the model was an appropriate representation for our data, we analyzed the deviance residuals.

These residuals should be roughly symmetrically distributed about zero with a standard deviation of 1. Furthermore, positive values correspond to satellites that "died too soon" compared to expected survival times. Conversely, negative values correspond to satellites that "lived too long". Very large or small values are outliers, which are poorly predicted by the model.

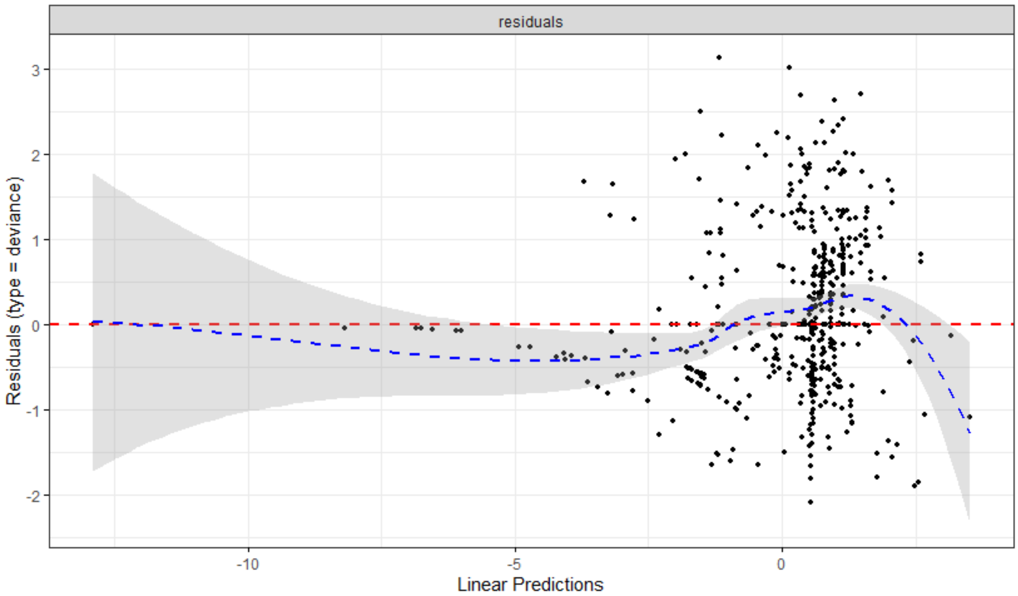


Figure 6: Deviance Residuals

Our model is capable of adequately representing our data since the residuals exhibit almost all the characteristics mentioned above.

4.3. PCA and Penalized Survival Analysis

Our dataset is affected by a high collinearity between the covariates, as can be seen from the covariance matrix reported in Figure 7.

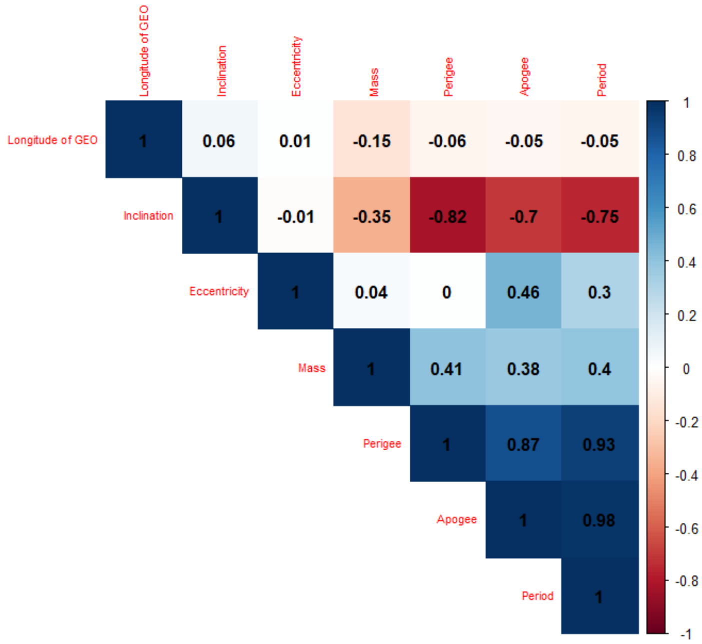


Figure 7: Correlation matrix

To address the issue and attempt to improve the model, we first tried PCA and subsequently a penalized Cox model.

First of all, in order to apply PCA, we considered only the numerical standardized covariates. The standardization process is advantageous in the presence of covariates measured on different scales, as in our case. This ensures that each covariate contributes equitably to the analysis. We chose to keep the first five principal components, so as to have a sufficiently high explained variance.

Furthermore, in order to include categorical variables, we also performed MCA (Multiple Correspondence Analysis), which can be viewed as the categorical data equivalent of Principal Component Analysis. This time we decided to take the first three components.

Applying the Cox model to these components, the only term that did not exhibit statistical significance was PC2. However, these findings were obtained without satisfying the proportional hazard assumptions. Therefore we were unable to proceed with this model. Moreover, we would have incurred a loss of interpretability.

At a later stage, we attempted a penalized approach. The Penalized Cox model is implemented in the `glmnet` library and is defined as follows:

$$\arg\max_{\beta} \log PL(\beta) - \alpha \left(r \sum_{j=1}^{P} |\beta_j| + \frac{1-r}{2} \sum_{j=1}^{P} \beta_j^2 \right)$$

where PL(β) is the partial likelihood function of the Cox model, $\beta_1, \ldots, \beta_p$ are the coefficients for p features, α is a hyper-parameter that controls the amount of shrinkage and r is the relative weight of the L^1 and L^2 penalty.

Despite several costly attempts to find r and α values that ensured a reduction in collinearity between variables and defining a model that satisfied the assumptions, we failed to find a valid one that was better of the model described above in terms of the goodness-of-fit. r values were manually changed, whereas α values were selected through cross-validation.

Both new approach, in addition to the issues already identified, show a higher AIC than that of the Stratified Cox model.

Model	AIC
Stratifed Cox model	2391.17
PCA + Cox model	3397.345
Penalized Cox model	2464.633

Table 6: AIC values

5. Prediction

At that point, our objective was to assess the superiority of our model compared to those currently in circulation and, furthermore, to comprehend how to derive the maximum benefit from our findings.

5.1. The `Coxed` Library

With this aim in mind, we found the predicted lifespan of the satellites using an R package called `coxed`. This package allows researchers to calculate duration-based quantities from Cox model results, such as the expected duration (or survival time) given covariate values.

In addition, no standard method exists for simulating durations directly from the Cox model's data generating process because it does not assume a distributional form for the baseline hazard function. The coxed package also contains functions to simulate general duration data that does not rely on an assumption of any particular parametric hazard function: the NPSF and the GAM method. As suggested in the associated paper [3], we decided to use the second approach since it produces better measures of uncertainty and its confidence intervals are closer to covering at the nominal level compared to NPSF.

The method chosen employs a generalized additive model (GAM) to map the model's estimated linear predictor values to duration times and proceeds according to five steps. First, it uses coefficient estimates from the Cox model. Then the method computes expected values of risk for each observation by matrix multiplying the covariates by the estimated coefficients from the model, then exponentiating the result. This creates the exponentiated linear predictor (ELP). Then the observations are ranked from smallest to largest according to their values of the ELP. This ranking is interpreted as the expected order of failure; the larger the value of the ELP, the sooner the model expects that observation to fail, relative to the other observations. The next step is to connect the model's expected risk for each observation (ELP) to duration time (the observed durations).

This package uses a GAM to model the observed durations as a function of the linear predictor ranks generated in the previous step. More specifically, the method utilizes a cubic regression spline to draw a smoothed line summarizing the bivariate relationship between the observed durations and the ranks. The GAM fit can be used directly to compute expected durations, given the covariates, for each observation in the data.

As examples, we obtained these expected durations for some satellites in the test set:

Satellite Name	Predicted lifetime	Effective Lifetime	Expected lifetime
Starlink-2216	3.338002	2.48	4
SpaceBEE-NZ1	3.061470	2.57	2
Iridium Next 123	7.837608	6.40	15

Table 7: Prediction of the lifetime

5.2. Boostrap Intervals

To produce estimates of uncertainty, the GAM approach repeats this process many times via bootstrapping. The method generates bootstrap samples of the data and re-estimates the Cox model coefficients on each bootstrap sample. At each iteration, this produces a new vector of actual durations and a new ranking of ELP values, which are then used to fit a new GAM. This results in a distribution of expected durations for each independent variable profile and a distribution of the marginal effect. These distributions can be used to produce standard errors and confidence intervals for the estimates. Importantly, by bootstrapping the entire process, this step incorporates the uncertainty from the Cox model estimation and the uncertainty from the GAM. This process is mostly automated in the `coxed` package.

For the same satellites in the table 7, the obtained bootstrap intervals are:

Satellite Name	Bootstrap se	lb	up
Starlink-2216	0.6625634	2.039402	4.636603
SpaceBEE-NZ1	0.8194228	1.455431	4.667509
Iridium Next 123	1.1727706	5.539020	10.136196

Table 8: Bootstrap CI for Predicted lifetime

6. Is Our Model Better Than the Existing One?

6.1. Nonparametric Test

We wanted to ensure that our model markedly deviates from the existing ones utilized by space agencies.
More specifically, we conducted tests to determine whether there was a disparity in the distribution between the lifespan predicted by our model and the expected lifetime estimated by space agencies.
Given that we were clearly working in a paired data framework, we opted for a Two-Sample Paired Univariate Permutation Test.

Let Y_1 be the random variable that denotes the lifespans of the satellites that our model has predicted and Y_2 the random variable that denotes the lifespans expected by space agencies.

The Hypothesis test is:

$$H_0 \ : \ Y_1 \overset{d}{=} Y_2 \ vs \ H_1 \ : \ Y_1 \overset{d}{\neq} Y_2$$

Under H_0 exchangeability is just between and within pairs (e.g., if we want to preserve likelihood, pairs cannot be splitted up).

The test statistics that we used is $T = |\bar{Y}_1 - \bar{Y}_2|$ and the resulted p-value of the test is zero, therefore the two models are significantly different at any confidence level.

It's to be noted that difference between the expected lifetime and the predicted one, doesn't verify the assumption of Gaussianity, in fact the Shapiro-Wilk normality test gives a p-value of 1.439×10^{-5}. This fact confirms how suitable it is for our case to work in a framework of Nonparametric statistic.

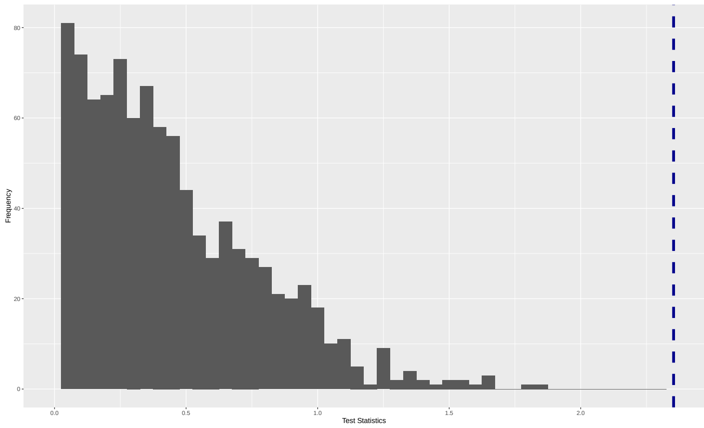


Figure 8: Histogram of the test statistics

We had reason to believe that our model was not just different, but it was better.
To prove this, we compared the Mean Square Error of our predicted lifespan with respect to the effective lifetime of retired satellites with the MSE of the predictive models used by space agencies.

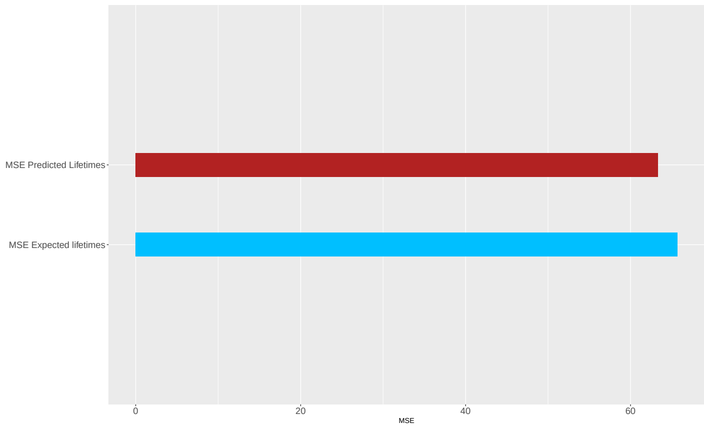


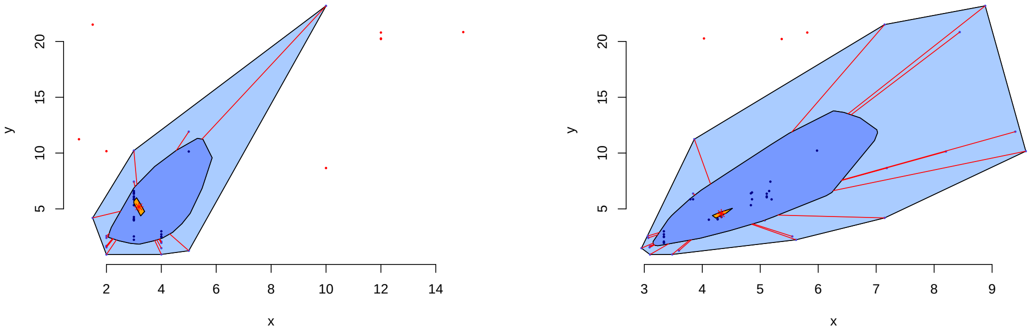
Figure 9: Comparison of MSEs

6.2. Outliers Analysis

To further evaluate the goodness of our model, we decided to analyze the presence of outliers in two different datasets as defined below:

- The first dataset contains the values of the lifetime predicted by our model and the actual lifetime of retired satellites present in the test set.
- The second dataset contains the values of expected lifetime provided by the UCS database and the actual lifetime of retired satellites present in the test set.

It should be noted that in the test set there are only 50 observations if only retired satellites are considered; therefore, the results below must be quantified with respect to the size of the two dataset.
In the event that the second dataset shows a greater number of outliers compared to the first, this confirms a greater accuracy of our model compared to the one used up to now.
Through the use of Bagplot, we found that our model was characterized by only 3 outliers compared to the 8 present in the second dataset. The two bagplots are shown below.



(a) Bagplot condidering existing model (b) Bagplot considering our estimates

Figure 10: Bagplot

We can therefore conclude that our model is reliable and more accurate than the one currently available on the market.

7. Where to Establish Our Business

We were at a point where we needed to assess the validity of our marketing strategy. The question was, should we choose the market where to start our business based on the continent where there is the greatest deviation between our predictive model and the actual models used by space agencies?
In order to answer to this question we performed a Permutational One-Way Anova on the difference between our predicted lifetime and the expected lifetime, where we tested the continent as an effect.

Suppose g to be the number of continents and n_i the number of observations relative to the continent i, then:

$$Predicted_{ij} - Expected_{ij} = \mu + continent.effect_i + \epsilon_{ij}; \quad i = 1, \ldots, g; \; j = 1, \ldots, n_i$$

$$H_0 : \; continent.effect_i = 0 \; \forall i \quad vs \quad \exists i \; s.t. \; continent.effect_i \neq 0$$

We used as a test statistics the Fisher's statistics:

$$T_F = \frac{\sum_{i=1}^{g} (\bar{Y}_i - \bar{Y})^2/(g-1)}{\sum_{i=1}^{g} \sum_{j=1}^{n_j} (Y_{ij} - \bar{Y})^2/(n_i - g)}$$

The p-value of the ANOVA is 0.039, therefore there is reason to believe that the continent is a significant effect in the determination of the difference between our model and the one already used by space agencies.

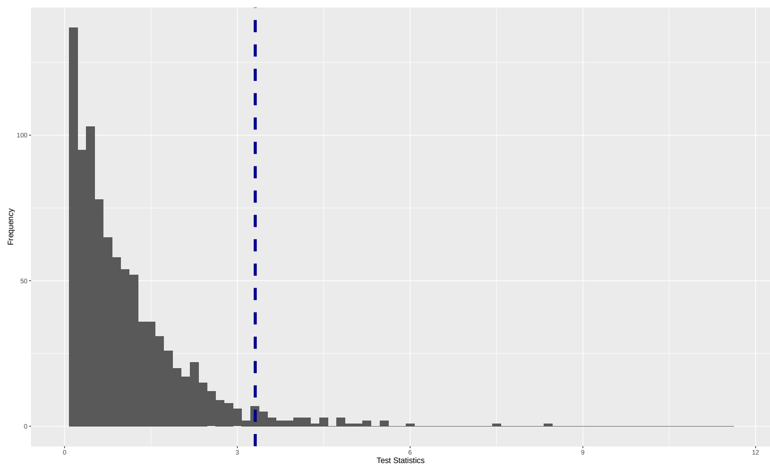


Figure 11: Histogram of the test statistics

We performed bootstrap reverse percentile confidence intervals in order to look for the continent that deviates the most from our model:

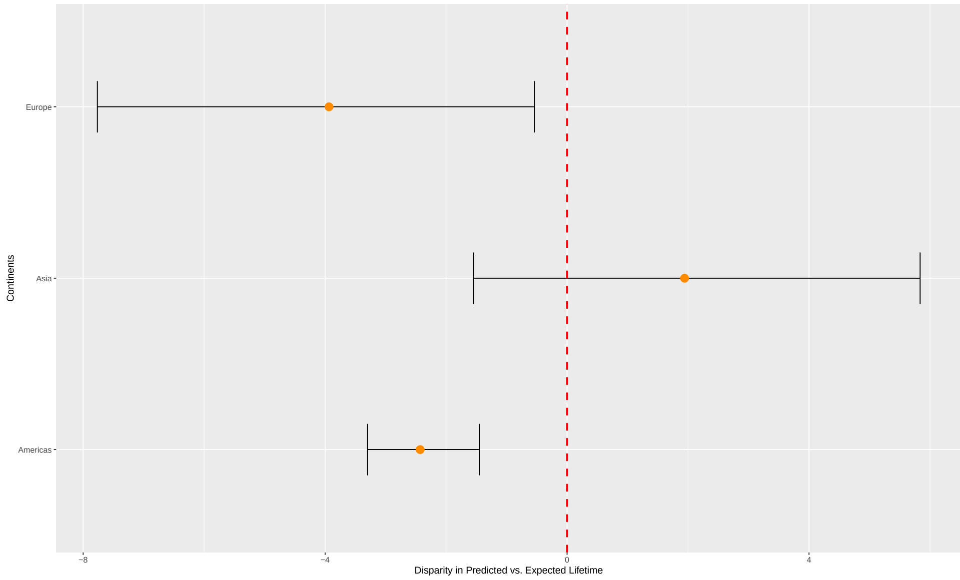


Figure 12: Bootstrap Intervals

These confidence intervals reveal that the difference between our predicted and expected lifetimes is not statistically significant for Asiatic satellites at a 95% confidence level. However, the difference is statistically significant for Europe and America. Subsequently, our inquiry aimed to identify the continent where our predictive model diverges the most from the expected, so to focus our resources in promoting our start-up in that market.
With this aim, we also determined the significance of the difference of our response between Americas and Europe. Once again, utilizing a bootstrap confidence interval, we observed that the difference is not statistically significant (CI: [-3.4882174, 0.4899549]). Subsequently, we made the strategic decision to focus on the market with the highest number of satellites, namely the American one.

8. Preventive Maintenance

Following our assessment that the American market is where we could best leverage our work and maximize profits, we have decided to focus on active American satellites for the continuation of our analysis. Our objective was to create a decision-making strategy based on our model to provide guidance to our stakeholders for preventive maintenance, thereby avoiding the unfortunate scenario of losing their satellites, which would result in severe economic damages.
Once we obtained the survival curves predicted by our model for each American satellite, we decided to set the critical threshold for the satellites' survival probability at 10%. We chose this threshold to have a conservative estimate of the time when the satellite might cease to function, allowing maintenance agencies to act promptly. We then calculated the 0.1 quantile of each survival curve and consequently the expected date for maintenance, obtained by adding the quantile to the launch date.

By looking at the 0.1-quantile distribution, we observed the following:

- For some observations, the quantile was not defined as the survival curve did not drop below 0.1, meaning that according to our model, the satellite will never reach a critical survival threshold. In other words, if it were to fail or be dismissed, it would be due to human decisions or unforeseen events (e.g., collision with an asteroid or other space debris) and not due to mechanical/software problems, and therefore it would not require maintenance;

- For some other observations, the expected date for maintenance was before today's date, meaning that according to our model, maintenance would have been carried out well in advance. This would certainly have involved a cost, but the maintenance would have extended the satellite's life even further. Moreover, this cost would be less than the cost of completely losing a satellite, and when large sums of money are at stake, it's better to be conservative;

- For all other satellites, it was possible to find a valid date and consequently create the calendar shown below, which reports for each month of the next 22 years the number of satellites that should undergo maintenance.

For illustration purposes, we present the features of a few randomly selected American satellites from our dataset, as well as their survival curves.

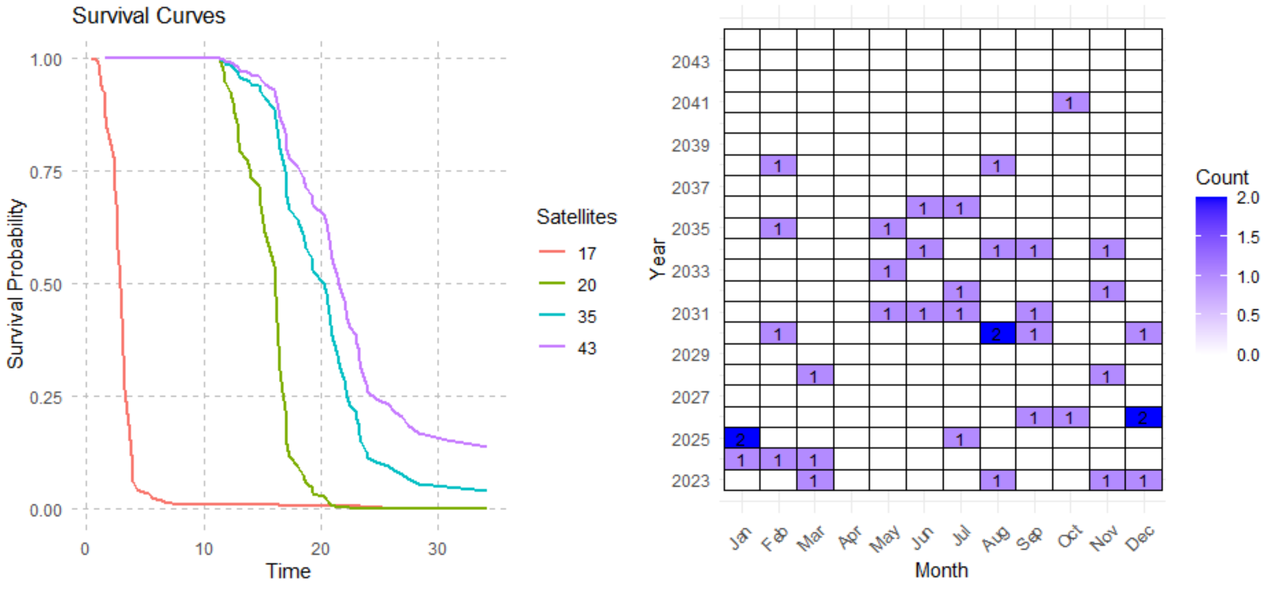


Figure 13: Survival Curves and Maintenance Calendar

Satellite Name	Apogee	log.Eccentricity	Mass	Users	Purpose	Orbit	Maintenance Time	Maintenance Date
ORBCOMM FM-23	794	7.777793	2	Commercial	Communications	LEO	3.91	2002-08-20
AMC-4	35791	9.039694	6	Commercial	Communications	GEO	18.01	2017-11-11
Galaxy-3C	35788	9.955985	5054	Commercial	Communications	GEO	25.80	2028-03-27
Spaceway F2	35787	10.649132	6000	Commercial	Communications	GEO	NA	NA

Table 9: Description of Satellite Data

9. Conclusions

In conclusion, the analysis we conducted reveals that the existing models do not offer a high degree of precision in estimating the lifespan of a satellite. This lack of precision poses a significant challenge for space agencies. However, the model we have proposed and developed in this project offers a more accurate alternative, ensuring fewer satellite losses and a more precise maintenance schedule.
This not only results in reduced costs for space agencies but also garners favorable evaluations from both the government and public opinion. Moreover, our model's ability to reduce costs and increase efficiency in satellite maintenance could have broader implications for technological advancements and sustainability, aspects highly valued by both the public opinion and public governments.

Our strategic focus on the American market, which holds the most promise for our business, has allowed us to maximize the impact of our solution. By setting a conservative critical survival threshold for satellites, we have provided a robust strategy for maintenance scheduling. Our model has demonstrated its value in predicting the lifespan of satellites, with some never reaching a critical survival threshold and others benefiting from early maintenance.

While our model offers a more accurate alternative for estimating satellite lifetimes, it is important to acknowledge its potential limitations. Predicting the lifespan of a satellite is inherently complex due to the multitude of factors that can cause a satellite to cease operation. For instance, strategic political decisions can lead to the premature decommissioning of a satellite. Similarly, unforeseen events such as collisions with space debris can abruptly end a satellite's operation. Furthermore, factors such as technological obsolescence, changes in mission objectives, or budgetary constraints can also influence a satellite's operational lifespan.
Therefore, while our model strives to account for as many variables as possible, the unpredictable nature of these factors may introduce some degree of uncertainty in our predictions.

10. Further Developments

Several avenues for a further development remain, presenting opportunities to enhance the depth and precision of our analysis:

- We could conduct an in-depth analysis of the reasons why each satellite is decommissioned. In the event of a malfunction, a risk and failure analysis could be carried out to propose a more precise schedule based on the type of failure;

- Following a new data collection, thereby obtaining a more balanced dataset, it would be possible to investigate which sector proves to be more profitable for our model. Specifically, we could identify the sector where there is a greater disparity between our predicted values and those obtained from the existing model;

- Another extension could involve considering an additional "Maintenance" variable, which indicates whether a satellite has undergone maintenance. This would enable us to provide a second maintenance date and repurpose the model accordingly.

These extensions could potentially enhance the precision and applicability of our predictive model in the dynamic field of satellite maintenance

References

[1] G. Fox, R. Salazar, H. Habib-Agahi, G. Dubos, *A Satellite Mortality Study to Support Space Systems Lifetime Prediction*, Conference: 2013 IEEE Aerospace Conference, California Institute of Technology, 2013, pp. 1–11.

[2] S. Grant, Y.Q. Chen, S. May, *Performance of goodness-of-fit tests for the Cox proportional hazards model with time-varying covariates.*, British Journal of Cancer, UK, 2003, pp. 1–7.

[3] Kropko, J. and Harden, J. J. (2020). Beyond the Hazard Ratio: Generating Expected Durations from the Cox Proportional Hazards Model. *British Journal of Political Science*, 50(1), 303–320. DOI: 10.1017/S000712341700045X.

[4] UCS Satellite Database Version 1.1. (2023). UCS Satellite Database. Union of Concerned Scientists. `www.ucsusa.org/satellite_database`

[5] Satellite Catalog (2023). `https://celestrak.org/satcat/search.php`.