Capstone project: Surviving the App-pocalypse

Objective:

The goal of this capstone project is to analyze app churn on the Google Play Store and build a predictive model using Survival Analysis and Cox Proportional Hazards Regression. We will identify the factors that contribute to app churn and estimate the survival probability of apps on the Google Play Store.

Background:

The Google Play Store is a massive marketplace for Android apps, with millions of applications available for download. App developers are continuously trying to engage users and retain them for as long as possible. However, app churn, defined as apps no longer being updated is an issue for the Play Store with lower-quality apps being available for download.

Understanding the factors that contribute to app churn and predicting the likelihood of an app being abandoned can help developers make data-driven decisions to improve their apps and marketing strategies.

Data:

The dataset for this project is provided, containing information on various apps available on the Google Play Store, and it included the following variables:

- App Name
- Category
- Rating
- Number of Reviews
- Size
- Number of Installs
- Type (Free or Paid)
- Price
- Content Rating
- Genres
- Last Updated
- Current Version
- Android Version

Tasks:

Data Preprocessing: Clean and preprocess the dataset, handling missing values and transforming categorical variables into numerical ones as needed.

Exploratory Data Analysis: Perform an exploratory analysis of the dataset to understand the distribution of variables and their relationship with app churn.

Kaplan-Meier Estimation: Estimate the survival probabilities of Free vs Paid apps using the Kaplan-Meier method.

Cox Proportional Hazards Regression: Build a Cox Proportional Hazards model to identify the factors that contribute to app churn.

Model Evaluation: Evaluate the performance of the Cox Proportional Hazards model using the Concordance Index.

Interpretation and Recommendations: Analyze the results and provide actionable insights and recommendations for app developers to reduce app churn and increase user retention.

Deliverables:

A well-documented Python Notebook or Python script containing the code and analysis for each task.