**Project Report – Cashflow Forecast 2021 to 2031**

GitHub URL - <https://github.com/magsmurphy/UCDPA_MargaretMurphy>

**Abstract**

The project uses Python to merge different data sets to project the future cashflow from premium seat holders in Croke Park Stadium from 2021 to 2031. It also looks at how credits would be used in lieu of payment in future years.

**Introduction**

There was a strong business need within Croke Park stadium to be able to project future cashflows for this area of revenue. This is the largest revenue stream in the company and a lot of the strategic planning would revolve around this revenue stream. The company was seriously affected by Covid 19 restrictions. As seat holders had no access to their seats for 2019 & 2020 they were given a credit on their account. This credit can be used against future renewals of contracts. It is vital for the company to predict when these credits are used so that they can see does it interfere with cashflow.

**Dataset**

The dataset was made up of 3 csv files from the company’s ticketing system. These showed how much credit each customer has in their accounts currently (‘Covid Credit.csv’), price plans showing when payments are due (‘Price Plans.csv’) and contracts (“Contracts.csv”) showing a list of current contracts their end date and what price code they are currently on.

There is also an excel file containing 3 sheets showing current debtors due, upfront sales due and future scheduled payments due on the contracts. (‘Debtors.xlsx’). This excel sheet was made up of different reports from our accounting software.

**Implementation Process**

The 3 csv’s were downloaded from the ticketing system and stored in a folder and committed to github repository. The excel file was then saved onto the same folder and again committed to the Github repository.

A project was then created on Pycharm called UCDPA. Pycharm was then used to clean and manipulate the data into the correct format for plotting on to a stacked bar chart.

I imported panda in PyCharm to allow me to read the excel sheets and create dataframes. Each sheet was read, column names changed and filtered into the dates needed. I then used the concat function to merge the 3 sheets together and called it deb

I then imported and read the csv’s ‘Price Plans’ and ’Contracts’. The two dataframes were merged in an outer\_join so that the payment amounts of each contract code could be added to the individual contracts. I then tidied the data by adding filling nan values with 0.

Created values for each year based on price plan multiplied by number of seats. Converted the end date to datetime. I then created payment due dates based on the end dates of the contracts & the payments due per pricing plan. Created separate dataframes for each year. I renamed columns so that the column names would be the same as the deb dataframe.

I then used concat to merge the 11 dataframes together. This now produced a dataframe with a list of every payment due between 2021 and 2031 showing the date and account number. This was named contracts. Any zero values in amounts were removed.

Covid credit.csv was then read and changed to a dataframe, cov1. This was tidied up by removing zero values.

Contracts dataframe was further tidied up by converting the column ‘acct\_id’ to an integer. The year column was added to the dataframe based on payment date. The dataframe was then grouped by acct id and year. This was then unstacked to allow the acct id be the index column. This was used as the original data frame had no unique columns in which to index the data. The process above allowed me to index on acct\_id. This was then sliced to only show data from years 2021 to 2031 using iloc called df.

Acct id was set as index for cov1 and this was merged to df using pd.concat function. Columns were then created for each year showing the amount used to pay the scheduled payment due by covid credit or payment. The covid credit balance was then reduced accordingly. Np.select function was used to create these columns based on conditions and choices.

The total sum of each column was created as a dataframe. Two columns were added to each index showing the years and category each index related to.

I then imported matplotlib and seaborn. A stacked barplot was then created based on the category type payments & covid credit for each year.

**Results**

A stacked barchart was created showing the cashflow and covid credits used forecasted for the years 2021 to 2031 inclusive.

Chart, bar chart

Description automatically generated

**Insights**

* 2025 will see the highest value of cashflow of €13 million so any future capital projects should be planned for then.
* 2025 will also see the highest value of covid credits used of €3 million.
* There will be a high value of covid credits used at the end of 2021. (€2.5 million)
* By 2028 the vast majority of accounts will have used their covid credit against future renewals so all payments will be cash based
* 2023 will be the lowest injection of cash with only €3 million forecast in cash receipts

**Machine Learning**

Machine learning could be used in future on the contract data. The forecast assumes that everyone will buy the same price plan as they had previously for the same term. As we see actual results coming in over the years this could be used as a test set along with actuals from previous years to predict whether people decide to renew or if they choose a different plan.