Ross Coleman

rosscolemandrc@gmail.com

IRISH CAR MARKET ANALYSIS

Abstract

The real-world business scenario I used in my project was that of a prospective car salesman analyzing data to gain insight about the current car market conditions in Ireland. Using Python, I investigated the demand for New Cars VS Used Cars in Ireland over the last 5 years, found out what where the most popular cars registered in Ireland in 2021 and also the 2021 market breakdown in terms of High-End VS Mid-Range cars. I plotted my results and upon generating my graphs I was able to extrapolate valuable insights that would be valuable for any car dealer in the country.

https://github.com/rosscoleman99/UCDPA\_RossColeman

Contents

[Introduction 2](#_Toc96028413)

[Dataset 2](#_Toc96028414)

[Implementation Process 2](#_Toc96028415)

[Step 1: New Car 1st Time Registrations VS Used Car 1st Time Registrations from 2017 to 2021 2](#_Toc96028416)

[Step 2: Top Selling New Cars 2021 3](#_Toc96028417)

[Step 3: Market Demand of High-End Cars VS Mid-Range Cars 3](#_Toc96028418)

[Results 4](#_Toc96028419)

[Plot 1: 2017 to 2021 Car Registrations 4](#_Toc96028420)

[Plot 2: Top 5 Most Popular Cars in 2021 5](#_Toc96028421)

[Plot 3: 2021 Market Demand of High-End Cars VS Mid-Range Cars 5](#_Toc96028422)

[Insights 6](#_Toc96028423)

[Plot 1: 2017 to 2021 Car Registrations 6](#_Toc96028424)

[Plot 2: Top 5 Most Popular Cars in 2021 6](#_Toc96028425)

[Plot 3: 2021 Market Demand of High-End Cars VS Mid-Range Cars 6](#_Toc96028426)

# Introduction

In my real-world business scenario, a prospective car salesman turns to Python to get a gauge of current market conditions in the hope to gain insights. He wants to find out whether he should focus on importing cars from abroad for example the UK and selling them on in Ireland or would it be more lucrative to get a dealership agency. He also wants to learn if there are certain periods, he should hold more stock for or does the demand remain relatively constant. He is interested in seeing what the most popular cars are registered in Ireland in 2021 to act as a guideline if he decides to import from abroad. Finally, he wants to see a breakdown of the 2021 market segmentation between high end cars registered versus mid-range cars registered to decide if he should focus in on a particular segment. I chose to use this scenario as I have an interest in cars and I was intrigued to see what I could learn using the available data from the Central Statistics Office. Using the skills which I have learnt through completing this course, I felt that I could efficiently extrapolate the data required and illustrate it in a clear way thus providing insights.

# Dataset

I got my chosen datasets from the Central Statistics Office’s website, data.cos.ie. The Central Statistics Office has a large array of data available in relation to economic, social and general activities and conditions in Ireland so I knew there would be my best bet to get data about cars being registered in the country. I downloaded my 2 CSVs files from dataset TEA01 one detailing “New Private Cars” registered for the first time whereas the other detailed “Secondhand Private Cars” registered for the first time.

For my API I intended to use the dataset TEM28 also from the Central Statistics Office’s website, data.cso.ie. The dataset was on “New Private Cars Licensed for the First Time” and this detailed what makes and models of cars were registered for the first time in the given year. I was having difficulty importing the API and was unable to contact my lecturer, so I just went ahead and imported it through a CSV file.

# Implementation Process

I’m going to breakdown my coding process into three different steps describing what I did in my Jupiter Notebook which you will find in the ZIP file or in my GitHub repository.

My first step was to compare new car first time registrations for 2021 with used car first time registrations for 2021. My next step was to find the top selling new cars of 2021. My final step was to breakdown the market segments in terms of high-end cars registered in 2021 versus mid-range cars registered in 2021.

## Step 1: New Car 1st Time Registrations VS Used Car 1st Time Registrations from 2017 to 2021

I imported my 2 CSVs into my notebook through pd.read\_csv() to convert them into a Pandas DataFrame. I then explored both using .head(), .tail(), .info(), .columns and .shape in order to get a scope on the data. I also checked for NaNs but fortunately there was none. I then merged the two DataFrames on the “Month” column and added the suffixes “\_new” and “\_used” to identify which DataFrame each column had originated from. I used pd.merge\_ordered as I was ordering time series data. I then saved this DataFrame as private\_cars\_1st\_time\_registrations. I then disaggregated the DataFrame only get data relating to the months from 2017 to 2021. I set the index to “Month” and sliced [“2017M01”:“2021M12”]. Sub setting using indexes is much simpler than writing the code for .isin so I carried on this pattern throughout my notebook. I then saved this DataFrame as car\_registrations\_from\_2017\_to\_2021. At this stage the DataFrame needed tidying up. I dropped the unnecessary columns using .drop() and renamed the remaining columns with .rename(). I then used a for loop to change the row labels to have the same format as international standard of dates. This makes it more reader friendly in my opinion. Finally using MatPlotlib, I plotted the DataFrame using a line plot.

## Step 2: Top Selling New Cars 2021

I intended to import this dataset as an API using requests.get() and pd.DataFrame(cars.json()).head() but it did not import correctly. To proceed I instead downloaded it as a CSV and imported it into my notebook through pd.read\_csv() to convert it to a Pandas DataFrame. Once again, I explored it and checked for NaNs. This time the dataset did include some and I created a function to tell me what percentage of the dataset had NaNs to see if it would still give an accurate representation. Only 6.45% of values where NaNs, a quite small percentage so I proceeded and dropped them. From exploring this dataset, I noticed that the “Statistic” column had “New Private Cars Licensed for the First Time” and “New Private Cars Licensed for the First Time Rank”. I only wanted “New Private Cars Licensed for the First Time” so I set the index to “Statistic” and sliced .loc[:"New Private Cars Licensed for the First Time"]. To ensure I only had data for 2021 I set the index to “Year” and sliced using .loc["2021":]. I then reset the index and pulled out only the columns I want from the DataFrame. I then renamed these columns. Following that, I wanted to investigate the “Make and Model” column so I used .unique() to pull out the columns unique values. Upon doing that I noticed the unwanted "Other" and "All models" values and used the .drop() method to drop the rows containing these values. The DataFrame was now saved as cars2021 but the values in the “Registered” column were floats and I wanted integers as its not possible to rent .5 of a car. To change this I called cars2021["Registered"].astype(int) and then saved the updated cars2021 to cars\_registered\_2021. I sorted the DataFrame to display the top registered cars in descending order using .sort\_values(). I wanted to plot the top 5 most registered cars, so I called cars\_registered\_2021.head(5) and saved it as top\_5\_most\_popular\_cars\_for\_2021. I then set the index “Make and Model” and plotted a bar chart with MatPlotlib.

## Step 3: Market Demand of High-End Cars VS Mid-Range Cars

I renamed the cars\_registered\_2021 to car\_brands and used .slice[:] again to select only rows where a car had been sold. To extrapolate the car make form the data I called make = car\_brands["Make and Model"].str.split(" ", n = 1, expand = True) and added it into the DataFrame. I then inspected the unique values of this new column once again using the .unique() method. Following inspection, I created a dictionary to include what I thought were premium brands and called this dictionary premium\_dict. I used a dictionary over a list because dictionaries are faster when looking up elements and the order of the values did not matter. I then called np.where(car\_brands["Make"].isin(premium\_dict),"High-End", "Mid-Range") so that where the values of the “Make” column where found in the premium\_dict a new column titled “Classification” was populated with “High-End” in them rows and if the values of “Make” column where not found in the premium\_dict the rows of the “Classification” column where populated with "Mid-Range". Once again, I tidied up the DataFrame by reordering the columns using .reindex() and set the index to “Classification”. I wanted to zone in on quantities of registered cars in terms of their classification so called car\_brands.groupby("Classification", as\_index=False)["Registered"].sum(). I saved this DataFrame as market\_breakdown, set the index to “Classification” and plotted a bar chart with MatPlotlib.

# Results

My coding process produced 3 plots that I have included in this document but can also be found in the ZIP file and my GitHub repository.

## Chart, histogram Description automatically generatedPlot 1: 2017 to 2021 Car Registrations

Using MatPlotlib, I created a line plot graphing the 1st time registrations of both new and used cars for each month from the start of 2017 until the end of 2021. I used a line plot as I felt it would best showcase the fluctuations of both the new and used car’s registrations over the five years. I set the plot’s style to “seaborn-colorblind” due to the fact the style offers gridlines and because people who are colour blind can read it. 1 in 20 people are colourblind and “seaborn-colorblind” uses colours that they can see. The colour for the used cars was green originally but I thought appeared too similar to the blue, so I changed it to red to help aid the viewer distinguish between the two lines. I added a plot title, set axis labels and rotated the xticks to make the plot more descriptive. I also added a legend to explain which line plotted what. The blue line plots the 1st time registrations for new cars for each month over the last 5 years whereas the green line plots the 1st time registrations for used cars for each month over the last 5 years.

## Plot 2: Top 5 Most Popular Cars in 2021

Chart, bar chart

Description automatically generated

Using MatPlotlib I plotted the top 5 most popular cars registered for the first time in 2021 with a bar chart. I set the style to “seaborn-colorblind” again to avail of the gridlines. I plotted with a bar chart as I wanted to display the quantity of my numerical data in this case no. of cars registered for each group in my categorical data in this case the make and model of car**.** I set the colours of the bars to resemble the manufacturer’s brand colours; gold representing Hyundai gold, red representing Toyota red, blue representing Volkswagen blue and green to represent Skoda green thus making the chart more visually appealing and informative.

## Chart, bar chart Description automatically generatedPlot 3: 2021 Market Demand of High-End Cars VS Mid-Range Cars

In my final plot I once again used a bar chart to convey the information I had gathered in my DataFrame in relation to the 2021 Market Demand of High-End Cars VS Mid-Range Cars. I set the style to “seaborn-colorblind” again to utilise the gridlines the style offers. I had considered using a stacked bar to visualise this information but decided against it as I felt it was easier to compare the two categories in terms of there numerical values by leaving them side by side.

# Insights

I was able to collect many insights from the datasets through the plots I had generated.

## Plot 1: 2017 to 2021 Car Registrations

* There are more new cars registered than used cars.
* The registration of new cars tends to fluctuate a lot more than used cars especially when in the upcoming months to a new number plate being released so coming up to July or January.
* Registrations of new cars are at their highest at the start of a calendar year so if the prospective salesman intended on becoming a dealership agency he would need to be well stocked before the start of each year.
* The onset of the Covid 19 pandemic in April 2020 negatively affected both new and used car registrations.
* If he was to go down the route of selling used cars from abroad sales, demand remains relatively constant throughout.

## Plot 2: Top 5 Most Popular Cars in 2021

* The Hyundai Tucson was the most popular new car in Ireland in 2021.
* Toyota had two cars feature in the top 5 most popular cars registered in 2021, with the Toyota Corolla coming 2nd and Toyota Yaris coming 3rd.
* Almost twice as many Hyundai Tucsons where registered in 2021 compared with the 4th most popular ranked car for that year in the Volkswagen Tiguan.
* Through identifying the most popular cars if the car salesman has decided to sell second-hand imported cars, he should focus on acquiring these models to sell on.

## Plot 3: 2021 Market Demand of High-End Cars VS Mid-Range Cars

* High-end cars is a low volume market.
* A focus towards selling mid-range cars would provide a steadier trade thus more financially lucrative.