PROJECT REPORT

DATA422 - DATA WRANGLING

Dissecting Catch and Export Patterns of New Zealand Fish Species

Group Members: Yamika Gandhi, Rajendra Pandey, Sharon Dwilif Kumar, Haritha Parthiban

INTRODUCTION

New Zealand is a significant player in the worldwide seafood sector because of its well-known abundance and diversity of marine resources. The fishing industry in New Zealand offers a vast range of seafood items that are suitable for both local consumption and exportation abroad and it also plays a key role in the country's economy and cultural identity.

This dataset, which focuses on seafood exports from New Zealand provides insights into the catch amount and values of fish sent to different nations illuminating the dynamics of the fishing sector's global engagement in New Zealand. With an area of more than four million square kilometres in the ocean, New Zealand has one of the biggest Exclusive Economic Zones (EEZ) in the world (New Zealand Ministry for Culture and Heritage Te Manatu Taonga). Numerous fish species such as orange roughy, hoki, and snapper can be found in this vast marine region. Because of this, fishing has always been an important part of New Zealand's economy. This report dives into the process of wrangling extensive datasets related to catch volumes and exported fish species to uncover patterns and insights. With the use of R, we process, transform and enrich the data aiming to illuminate the dynamics between catch volumes, export values, and potential external influencers such as sea surface temperature. Through this, we aim to present a cleaned and well-organised dataset that paves the way for insightful investigations into New Zealand's marine resources.

Background:

An important part of New Zealand's economy is the fishing industry. The GDP, exports, and employment of the country are all greatly impacted by the seafood sector. In coastal communities, fishing, aquaculture, and related sectors provide jobs and revenue making them an essential part of the nation's economy.

The management of its fisheries is known to be strict and grounded in science in New Zealand. Strict laws and quotas have been put in place by the nation to guarantee the sustainability of its fisheries. A key component of New Zealand's strategy is the Quota Management System (QMS), which assigns catch limits to certain species in order to prevent overfishing. The QMS is evaluated and modified on a regular basis in accordance with scientific findings. Understanding the economics of fisheries requires consideration of market dynamics including price fluctuations, supply and demand trends, and consumer preferences.

New Zealand is committed to maintaining its distinctive ecosystems and safeguarding its maritime biodiversity. The data that provides insights into these dynamics often comes from varied sources, with inconsistencies and gaps. Different naming conventions for species, varying units of

measurement and the volume of data entries make the task of analysis challenging. By cleaning, and structuring these datasets, we aim to provide a foundational resource that can drive both academic research and policy decisions in the realm of New Zealand's fisheries.

RESEARCH QUESTION

IHow do variations in catch volumes of fish species affect export values and what factors influence it?

DATA

The New Zealand Ministry for Primary Industries has a comprehensive and detailed data repository regarding various fish species and their catch statistics. This information is made publicly available on their official portal, Fisheries New Zealand. The dataset offers insights into reported commercial catches, Total Allowable Commercial Catches (TACC), customary allowances, recreational allowances, and whether the fish species is a target or bycatch in the Quota Management System (QMS). We opted for this dataset due to the granularity of information provided. The information is only updated annually and it also provides historical data, enabling us to trace and analyze trends over time. Since there was no option to directly download the data, we scraped the data for the time period 2018 - 2022. In the report, we will be referring to it as the CATCH data.

This table has the following columns:

Species Code (SpeciesCode): Each fish species is uniquely identified by a code found in this column. It can be used as a guide to identify between various species in the dataset and is frequently employed for data organisation and retrieval.

Name (SpeciesName): The common or scientific name of the fish species is shown in this column. It gives the species denoted by the matching species code a label that is readable by humans.

Reported comm. catch (kg) (ReportedCatch): Information about the fish species' reported commercial catch is provided in this column. It stands for the total number of species that have been harvested for trade.

TACC (kg)(TACC): Total Allowable Commercial Catch is expressed as TACC (kg). The maximum permitted catch limit for the fish species is shown in this column in kilos. It is the total amount that commercial fishing operations are permitted to lawfully catch without endangering fish stocks.Custody

Cust. allow. (kg)(CustomaryAllowance): The customer's permitted catch, expressed in kilograms, is shown in this column. It could be used to describe a share of the overall catch allotted to particular clients or purchasers. This allocation is frequently made to guarantee that all parties involved receive a fair share of the catch.

Rec. allow. (kg)(RecreationalAllowance): Rec. allow. stands for Recreational Allowance. This column indicates the amount of the species allowed to be caught for recreational or non-commercial purposes, such as sport fishing. It is often a part of fisheries management to account for non-commercial fishing activities.

Target fishery(TargetFishery): This column indicates if the fish species is inadvertently taken while targeting other species, or if it is the major focus of the fishery (i.e., actively sought after). Values such as "Yes" for target species and "No" for bycatch species could be present.

Bycatch fishery (ByCatchFishery): This column shows if the species of fish is accidentally caught as bycatch in a fishery that mostly catches other species. Additionally, it might include values like "No" for target species and "Yes" for bycatch species.

In QMS YesNo (InQMS): Indicates if the fish species in question is subject to New Zealand's Quota Management System (QMS) regulations. Whereas a "No" implies that the species is not managed in this system, a "Yes" value shows that it is handled within the QMS framework.

Source 1: https://fs.fish.govt.nz/Page.aspx?pk=6&tk=97

Total food export from New Zealand to international markets in the year ended 2022 is $72.2 billion (NZ, 2023). Out of this total export, the total revenue generated from the seafood industry is $1.92 billion (Statistia, n.d.). It shows that the revenue generated from the seafood industry has remarkable contribution in New Zealand economics.

New Zealand seafood export data was retrieved from Seafood New Zealand. Individual yearly datasets from 2018 to 2022 were combined to make a single tibble. This dataset provides a breakdown of export country, species and production type and other numerical columns such as ExportWeight, TotalPrice and AverageUnitPrice.

The New Zealand seafood export data is analysed to find out various factors such as top importers of all species from New Zealand, top importers of specific species from New Zealand, and distribution of average unit prices over the years.

Source 2 - https://www.seafood.co.nz/publications/export-stats

These export stats are published each month. They are given in a Year-to-Date (YTD) format, which means the data is for that calendar year to date. If we take export data from December[required year], it will give us export data for that whole year. Clicking on the month provides us the excel sheets of the export data, by product type, by species, or by country. The latter 2 excel sheets have grouped and aggregated data and cannot be directly put in a dataframe. In the report, we will be referring to it as the EXPORT data.

SpeciesName (Seafood) - This column represents the name of the fish species being reported in the dataset. It contains information about the particular species of fish being tracked, such as the scientific or popular name of the species. Names like "Hoki,"Bluenose" or other species found in the waters of New Zealand, for instance, might be included.

ProductType: This column describes the type or category of seafood product. It specifies the form in which the fish product is exported.

Country (ExportCountry): This column indicates the destination country or market to which the fish products are exported. It specifies the location where the New Zealand seafood is being shipped for consumption or further distribution.

YTDNettWeight (Weight): The Year-to-Date (YTD) cumulative weight of the product type or fish species exported to the designated nation is shown in this column. A running sum of the exported weight from the start of the year to the present is shown by the YTD figures.

YTDValue (TotalPrice): This is the total value of fish exports to the designated country from year to date. It shows the entire monetary worth of fish items shipped from the beginning of the year to the present.

YTDNettUnitValue (AvgUnitPrice): The average unit value of the exported fish products is shown in this column. The YTDValue is divided by the YTDWeight to get the result. The unit value is a measurement of the average cost of the exported fish products per unit of weight, such as price per kilogram

MTDNettWeight: MTDNettWeight stands for Month-to-Date (MTD) net weight. It represents the cumulative net weight of fish species exported to the specified country within the current month. Like YTDNettWeight, the unit of measurement is typically kilograms (kg) or tons (t).

MTDValue: The month-to-date cumulative value of the fish exports for the designated nation for the current month is known as MTDValue. It shows the entire dollar amount of fish items exported for that particular month.

MTDNettUnitValue : This determines the average net unit value of fish products exported for the current month, same as YTDNettUnitValue. This figure represents the average price per unit of weight for the products exported in that month and is calculated by dividing MTDValue by MTDNettWeight.

METHODS AND TECHNIQUES

The first step was to scrape the CATCH data using the `rvest` package.

To extract data from the New Zealand fisheries website, we defined a base URL which served as the starting point to access data for different years. A function `scrape_and_store_data` was defined, which takes a list of years as input, and scrapes data for each year from the specified URLs. This function initializes an empty dictionary, `catch_data_dict`, to store the scraped data. The code then loops through each year in the year list using a for-loop. For each year, it constructs the URL by appending the year to the `base_url`. Then it scrapes the HTML content from the URL. The table is extracted using the CSS selectors and the `html_table()` function and the scraped table data is stored in the dictionary as `value`, and the year is used as the `key`. The function then returns this dictionary.

We used this function to scrape data for the years 2018 to 2022. Then we extracted individual CATCH tables for each year from this dictionary. Since the year information is missing in these dataframes, a `Year` column was added to each corresponding dataframe to maintain consistency. This data was then combined vertically using `bind_rows` function into a tibble called `catch_all_years`. The resulting tibble has 847 rows spanning across 10 columns, corresponding to 5 years.

In this data preprocessing phase, several modifications were implemented to refine the dataset. Firstly, we renamed the columns and eliminated whitespaces. The catch quantities needed to be transformed into double numeric type after removing commas using the `gsub()` function. Furthermore, the columns 'Target fishery', 'Bycatch fishery', and 'In QMS' were mutated to the 'factor' data type to represent their categorical nature. The 'Year' column was also recast as a 'factor' which would be helpful for year-wise analysis.

In the dataset, each yearly catch table consists of 170 rows, corresponding to 170 distinct `SpeciesCodes`. This indicates the absence of duplicates within this specific column for a given year. However, an observation of the `SpeciesNames` column revealed only 166 unique names. This discrepancy means that certain species names in our dataset are associated with more than one species code. It was important for us to identify and address these species names with multiple corresponding codes. To address this, we grouped the data by `SpeciesName` and then evaluated the number of unique `SpeciesCodes` associated with each name. This process identified instances where the count of unique codes was greater than one. We found that "Gemfish" for example, is represented by two different codes, "SKI" and "RSO". The ReportedCatch values for "RSO" in these years were also missing. How we dealt with this issue is detailed later in the report.

Next, we started working with our EXPORT data.

We imported the excel files into our Jupyter notebook, and then added a `'Year'` column to each dataset for clarity. Next, we removed unnecessary columns that described product types as this wasn't required for our current purpose. All the cleaned datasets from different years were then combined into one main dataset vertically using `bind_rows()` function. Some columns were dropped whereas others were renamed. There was a column called `SpeciesName` in the data, however, this name was misleading because it contained fish names in a manner that related to seafood. For example, "Octopus fresh or chilled" is not really a specie. It tells us about Octopus as a seafood. Therefore, we renamed this column to `Seafood`. The `'Country'` and `'Year'` columns were mutated to <factors> for easier analysis. We simplified the names of some countries, for example, changed "China, Peoples Republic Of" to just "China", so that they would be easier to plot on a map later.

Next, we examined the NA values in our data. Only the `AvgUnitPrice` column had some NA values. We looked at these rows. Interestingly, while the `'Weight'` column displayed a value of zero in these rows, the `'TotalPrice'` wasn't consistently zero. This was strange, because if no seafood was exported to a country (as indicated by the weight 0), logically, the total price should also be null. Since this data didn't make sense, we decided to drop these rows.

We also analyzed the distinct number of countries and seafood types in this dataset. There were 136 distinct countries and 285 seafood types present.

After the preprocessing, we started the process of joining the data.

To combine the data, we needed a common column which could act as the primary key. So we examined how many species names matched directly in the CATCH and EXPORT dataset. In order to ensure a case-insensitive match, these species names were converted to lowercase using the `tolower()` function. We found that there were only 24 matching species out of the 166 species in

CATCH and 285 seafood types in the EXPORT dataset. This was because the naming conventions for both datasets are different.

We looked further at the naming styles in both the datasets to figure out how to join them.

We discovered a few issues.

- There were naming inconsistencies even within the same species. For example, the CATCH data had “achovy” while the EXPORT data had “anchovies” .
- The species names were getting too specific. The CATCH data has several different species of crabs, like Paddle Crab, King Crab, Giant Spider Crab, and others. Meanwhile, the EXPORT data mentions crabs in a general manner, for example, “crab, frozen” or “crab cans or jars.” It’s impossible to identify which specie of crab is inside those cans.
- The same specie has many variants in the EXPORT data. For instance, the term "Octopus" was presented in six variants, such as 'Octopus fresh or chilled', 'Octopus frozen whole', and 'Octopus live', all of which essentially refer to the species "Octopus".
- The EXPORT data contained terms like “other molluscs” and “shrimps & prawns”. These are not really species. They are classes which contain many species. For example, we now know that “molluscs” include animals which usually have en external shell. This includes snails, mussels, oysters, clams, octopus, squid, paua, and other creatures.
- There were certain marine creatures that were present in the EXPORT data but not in the CATCH data, and vice versa.

We realised that we needed further understanding of the species and classes. We needed some kind of biological taxonomy information which was exhaustive and contained the species present in both CATCH and EXPORT datasets. Such a reference data could act as our middle table, and it would allow us to create a mapping between different naming conventions accurately. It would also provide a structure from classes to sub-classes and specific species. This could be a good way to reconcile species names by standardising them.

We found this TAXONOMY data on the Ministry of Primary Industries (MPI) website. This was good, because it was the same source from which we got the CATCH data. This meant that the TAXONOMY and the CATCH data had the same `SpeciesCodes`, and could be joined on them.

Again, we scraped this data. We created a custom function named `scrape_metadata`. This function navigates through the site's multiple tables recognized by unique CSS selectors. All the retrieved data tables were stored into a dictionary with the CSS selector as the key and the corresponding table as its value. We extracted the individual tables, assigned classes and subclasses where that information was not scraped, and used `bind_rows()` to join them all into a table called `taxonomy`. This table had 586 rows, each containing a unique species name with a unique `SpeciesCode`.

Initially, we considered creating `Class`, `Subclass` columns in the EXPORT data, but quickly realised that this wouldn’t work. We didn’t have all the information about species. For some, we only had the species name. For some, we had a general class like “snails”. For some, there were groups of fishes like “salmonidae”.

Finally, we decided to create a new column called `CommonName`, which would contain the most common names of species, which anybody can understand. This common name is not necessarily the specific specie name. It can be a class or a subclass also, depending on whichever name is the most popular to identify a creature. Sometimes, specific groups of fish are well-known, like salmon and tuna. Sometimes, specific species might also be well-known, like Blue Cod. And usually, the non-fish creatures are known by a general name like “sea cucumber”, without people knowing the specific species names.

Therefore, we decided it was best to create a `CommonName` column and assign names ourselves. These common names would also allow us to group creatures together. For example, all the species of crab mentioned above could be given the common name “crab”. We used regular expressions and the `grepl()` function for this.

The `CommonName` was assigned keeping in mind the species we had in the EXPORT data. For example, the EXPORT data had “salmon” and “salmonidae”. Seeing this, we created a `CommonName` called “Salmon, Trout & Other Salmonidae Fish”. Then we researched to find out which species come in the salmonidae family, and assigned this common name to those fish. This step took the longest time of all, and we took help from various sources, which we have mentioned in the “References” section.

For some species, such as “jack mackerel” and “leatherjacket”, we kept the species name in the `CommonName` instead of assigning a group of fish, because these species were direct matches and were clearly identifiable in the CATCH and EXPORT datasets.

The common names were then converted to factors for analysis purposes. These steps resulted in capturing 88 species, compared to the previous 24, which was an improvement.

We created a similar `CommonName` column in the EXPORT data. This dataset had 64 distinct `CommonNames`. There were 1749 rows where no category could be assigned, because the `Seafood` column didn’t provide species information. For example, “other fish packed” isn’t very useful to us. So we dropped the rows that had NA in the `` `CommonName` `` column.

After doing some further refinements in the naming conventions, the TAXONOMY and `catch_all_years` datasets were merged using a left join on the "`SpeciesCode`". Previously, we were facing an issue with certain codes – Green-lipped mussels had 2 different codes, “GLM” and “MSG”. The data corresponded to the “GLM” code, while the `taxonomy` table had the code “MSG” for these mussels. We changed this code in the `taxonomy` table so that when we joined, we could preserve the catch data for green lipped mussels. Perhaps this discrepancy happened because the code for green-lipped mussels was changed at some point of time.

There were a few `SpeciesCodes` that were in the CATCH data but not in the TAXONOMY data. To preserve the catch-data associated with these codes, we added 2 new rows in the `taxonomy` table with these codes and their common names.

Our new joined table was called `taxonomy_and_catch`.

These 2 tables have a one-to-many relationship. Every `SpeciesCode` corresponds to 5 different years. However, some of the `SpeciesCodes` had the same `` `CommonName` `` (for example, Giant Spider Crab, King Crab, Paddle Crab had now just become 'Crab'), so the `` `CommonName` `` was not

unique, and was repeated even for the same Years. So we grouped the tibble on `CommonName` and aggregated the `ReportedCatch` values. Before merging, we used `SpeciesCode` as the main identifier. After merging, we have a complex primary key, a combination of (`SpeciesCode, Year`).

This gave us the tibble `grouped_taxonomy_and_catch`.

In the EXPORT data, we did a similar aggregation as well, grouping on the unique identifier, which was a combination of (`CommonName, Year, ExportCountry`). We summed the `` `Weight` ``, `` `TotalPrice` ``, and `` `AvgUnitPrice` ``. This gave us `grouped_export`.

Finally, we outer-joined `grouped_taxonomy_and_catch` and `grouped_export` on (`CommonName,Year`). This gave us our final dataset, `relational_data`, which has 4995 rows and 8 variables.

RELATIONAL DATA MODEL

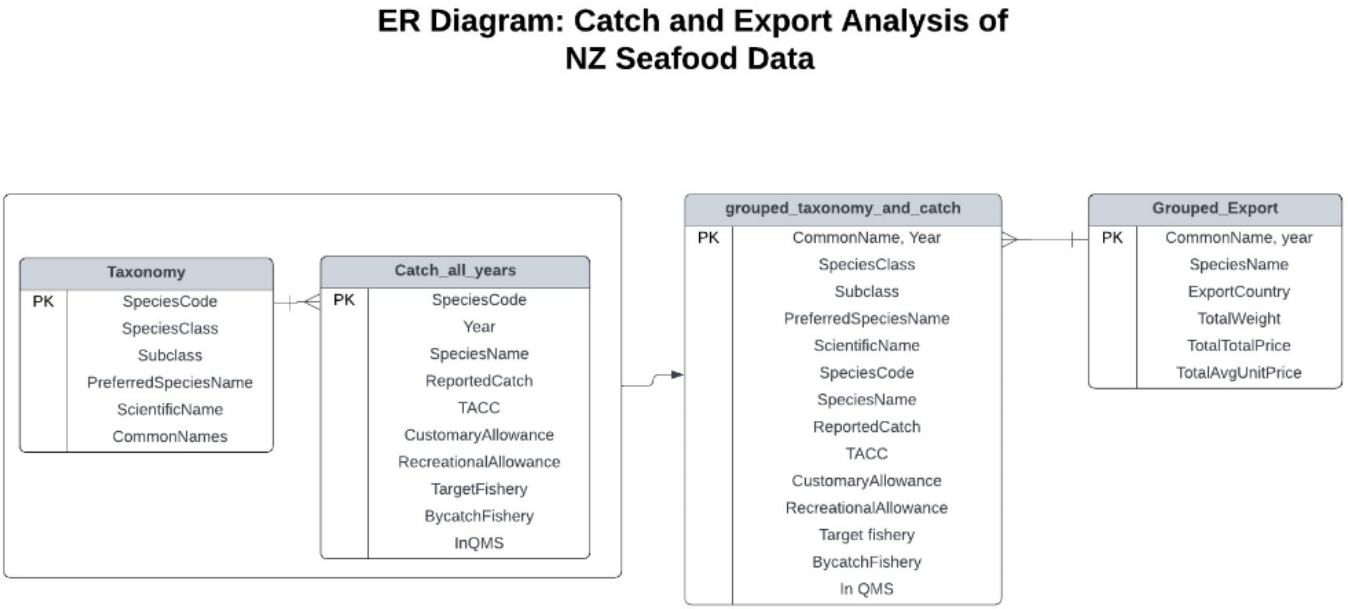

Figure1: Relational Data Model of Catch and Export Data

JULIA

INTRODUCTION

For this section, we manipulated 3 datasets to construct a smaller data model that would be useful to examine export values and catch volumes of New Zealand fish over a time frame of 2003 - 2013. External factors such as environmental and economic variables were taken into account as these often play a significant role in shaping the fishing industry. We decided to incorporate two external determinants: mean sea surface temperature (SST) and fuel prices to capture the trends in catch and export of seafood.

The mean sea surface temperature provides an ecological and environmental angle to the catch volumes and export values. Temperature variations can influence growth rates, spawning and behaviour of fish. For example, warmer waters might accelerate growth for some species while inhibiting it for others. This can impact the overall yield and quality of catches. On the other hand, incorporating fuel prices (mainly diesel) into the dataset gives an economical view to the catch export data. Since diesel prices play an essential role in operational costs of fishing vessels, transportation costs and profit margins, by merging this into the dataset, we can analyse the fluctuations in catch and export values and understand market dynamics over the years.

DATA

Three datasets were used to build the relational data model. The details are as follows:

1. Fish Monetary Stock Account:

The data ranges from 1996 - 2019 and includes 98 species. To customise, “All Species” was selected and this was exported as csv. The original table has the columns Species, year, variable and flags.
The “variable” had 5 unique values defined below:

- **Species**: Specifies the type of fish
- **Year**: The year for which the data is recorded
- **Asset value**: This represents the monetary value assigned to the particular fish species for the specified year. This column has missing values.
- **Catch**: The amount of the fish species caught in the given year. This column has missing values
- **Total allowable commercial catch**: This indicates the level (in kg) set by authorities on how much of the fish species can be commercially caught in a particular year. It's a measure implemented to ensure sustainable fishing practices.
- **Export value**: Denotes the financial worth of the fish species that was exported in a given year. This can provide insights into the international demand and profitability of the species.
- **Export quantity**: Represents the amount of the fish species that was exported.

The data is exported in a long format which is then pivoted into a wide table for analysis.

Source: https://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE8500

2. Mean Annual Sea Surface Temperature (SST):

The dataset includes the variations in sea surface temperature across 4 areas over a time period of 1993 - 2013. From the included report, we understand that the "mapped area" represents the overall spatial average sea surface temperature (SST) encompassing three specific regions: the eastern Tasman Sea, SubTropical Waters (STW), and Sub Antarctic Waters(SAW). For analysis purposes, we decided to choose the total area “MAP”.
The selected column descriptions are as below:

- **Year**: Ranges from 1993-2013
- **Maximum:** The highest recorded sea surface temperature for the specified area in the given year
- **Mean:** The average sea surface temperature for the area
- **Minimum:** The lowest recorded sea surface temperature for the specified area in the given year

The dataset along with the report is downloaded directly as a zip folder.

Source:https://data.mfe.govt.nz/table/52581-mean-annual-sea-surface-temperatures-19932013/

3. Fuel prices data:

The tables contain diesel, fuel oil, natural gas and petrol price data for New Zealand.

- **Fuel type:** The type of fuel being considered (e.g., Diesel, Petrol, etc.).
- **Category**: Commercial/Retail
- **Measure**: Real Average Price
- **Value**: Price presented in cents/litre
- **Value Unit:** NZD

This dataset tracks the real average prices of different types of fuels over various years. To find the data click the URL provided and click 'Energy prices'. This data was extracted and saved as csv:

- Sheet: 6 - Annual c per unit (real)

Source:https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-statistics/energy-prices/

METHOD

The fish monetary stock and temperature csv files were read into Julia and then converted into a DataFrame. The original fuel prices excel file had multiple sheets, so the required sheet was manually extracted to include data pertaining to commercial prices of diesel. This was then converted to a csv file for easier manipulation. This process was done on a separate .ipynb file and later joined.

During the data cleaning phase, we identified and addressed missing values to ensure data integrity. Unnecessary columns that weren't relevant for analysis were removed from the datasets. There were no duplicates identified in these datasets. Columns were renamed appropriately for clarity and consistency. Methods like 'dropmissing' were used to remove missing values in specific columns.

In the fish monetary stock account data, values under the 'Variable' column were identified using a unique function. This was important for the reshaping process. The unstack function was then used to pivot the data using 'Species' and 'Year' as the key columns. The columns 'Variable' and 'Value' were spread out to shape the data into its wide-format to get the catch, asset value, TACC, export value, export quantity columns. The resulting data frame provided a clearer view of each species' metrics by year. Eventually, we selected the aggregate values of all species for the data frame since this was more logical to compare with mean sea temperatures and fuel prices.

To provide a more comprehensive view of the catch-export dynamics, we incorporated data on the mean sea surface temperature (SST). This temperature dataset categorises measurements into four distinct sub-areas: the Eastern Tasman Sea, SubTropical Waters (STW), Sub Antarctic Waters (SAW), and the Mapped Area (MAP). For our analysis, we specifically selected the "MAP" category, as it offers a representation of the average sea surface temperature across regions.

Initially, we performed an inner join between the fish stock data and temperature data using the common "year" variable. Following that, we conducted another inner join, this time with the fuel prices dataset obtained from the "Fuel Prices.ipynb" file also using the "year" attribute as the key for merging.

We opted to focus on the aggregate data specifically choosing the "All species" rows to derive a better understanding of trends and factors influencing the fishing industry. This was done to reduce potential redundancy of having fuel prices and mean sea surface temperatures (SST) for individual species within the same year. This simplified the data structure as the data is available for 11 years only. However, it could be used to get an overall view of how the entire fishing sector responds to external factors such as fuel prices and SST.

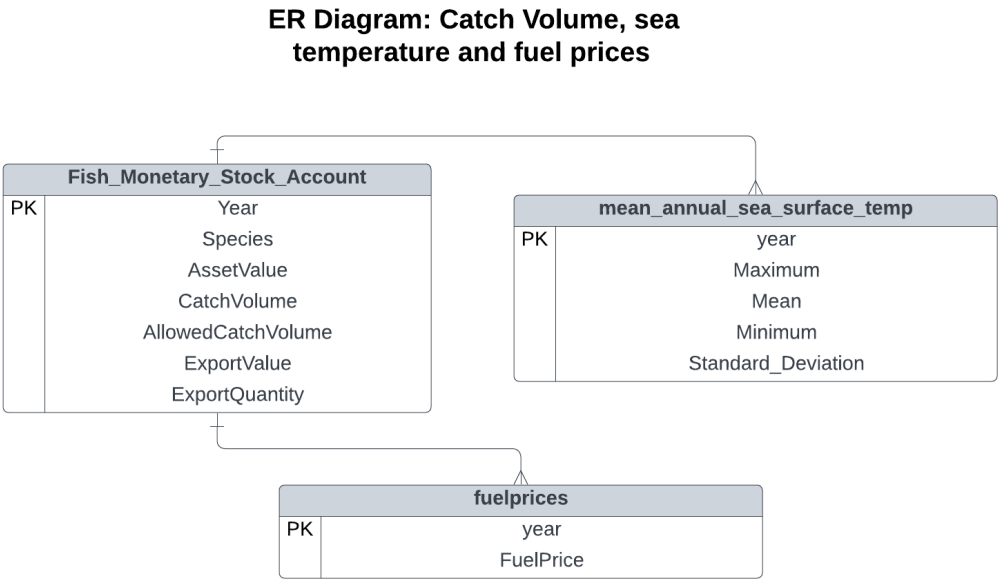

Figure 2: ER Diagram of Fish Monetary Stock Account, Mean SST and Fuel Prices

In order to integrate the datasets, an inner join was executed on the "Year" column, merging the catch-export and temperature data. Subsequently, another inner join was applied to incorporate the fuel prices data using the "year" column. The resulting dataset which is refined and organised spans 11 years (rows) and comprises 11 columns.

To visualise the patterns and trends present in the merged dataset, we utilised the VegaLite.jl package in Julia. The objective was to provide insights and a better understanding of how various catch, export, environmental and economic parameters correlate with each other over the years and how this data can be used for further analysis.

FINDINGS

1.1 Analysis of Catch and Export Data

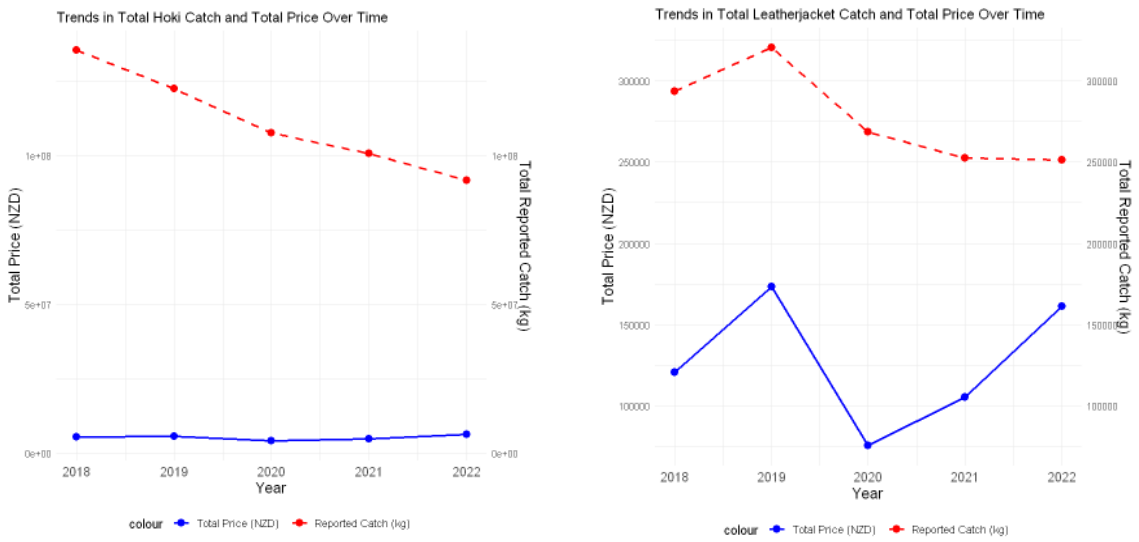

Figure 3:Analysis of Reported Catch and Total Price of Species over the Years

From the analysis of the total reported catch and total price of two species of fish (Hoki and Leatherjacket), we see that there is a correlation between these factors. Although the changes in price of Hoki seems low, there is a gradual rise in prices from 2020 onwards. This could be due to the reported catch reducing and hence affecting the supply-demand parameters. It could also be due to stricter regulations in fishing, reducing the TACC or other factors like COVID-19 leading to lesser fishing operations. The total price of Leatherjacket however has increased significantly since 2020 with decreased catch which could indicate that these were exported more at a higher price.

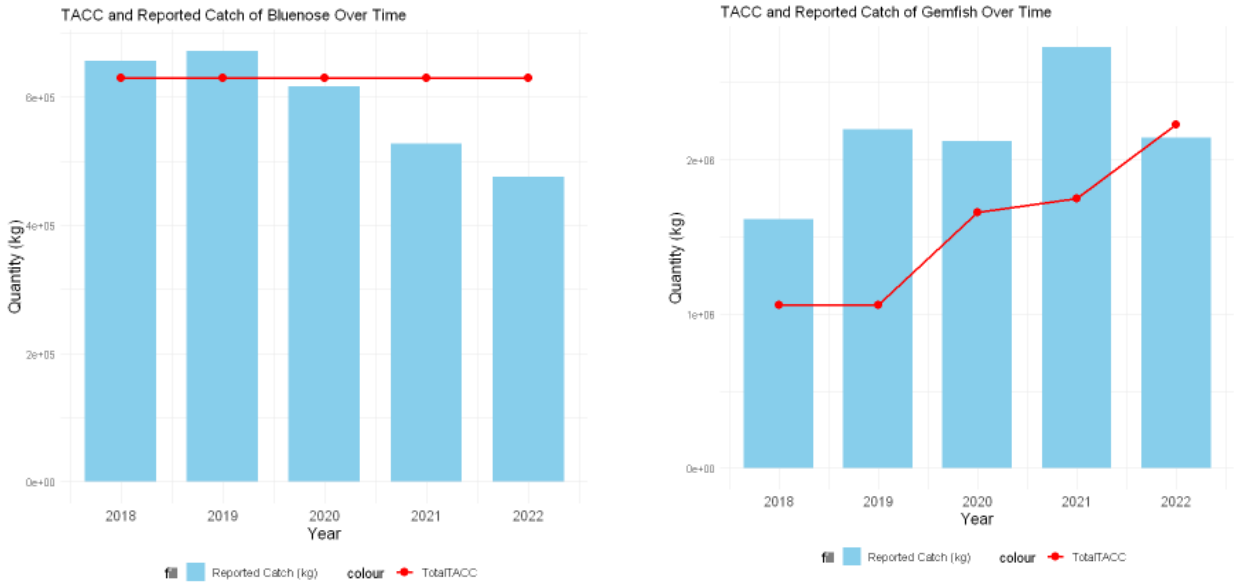

Figure 4: TACC vs Reported Catch of Species Over Time

An analysis of the total allowed catch and reported catch of species shows how these need to be closely monitored to ensure sustainable fishing practices. When the reported catch exceeds the Total Allowable Commercial Catch (TACC), it indicates potential overfishing which can lead to depletion of fish populations and long-term negative impacts on marine ecosystems. For example, in 2018 and 2019, Bluenose was overfished and from 2020 it was regulated. Similarly, Gemfish has been overfished between a period of 2018-2021. Continuous overfishing over such extended periods can lead to a significant decline in their populations making recovery challenging and time-consuming.

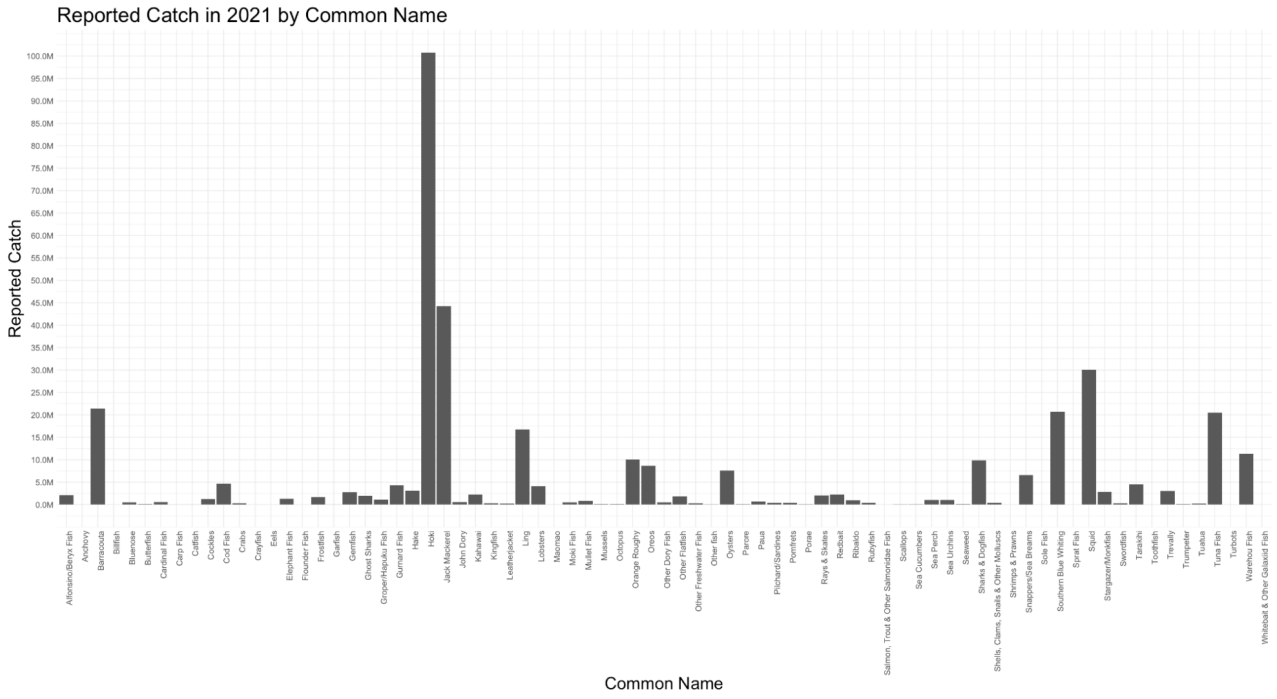

Figure 5: Reported Catch in 2021

The most caught fish in 2021 was Hoki, followed by Jack mackerel, and various Squid species.

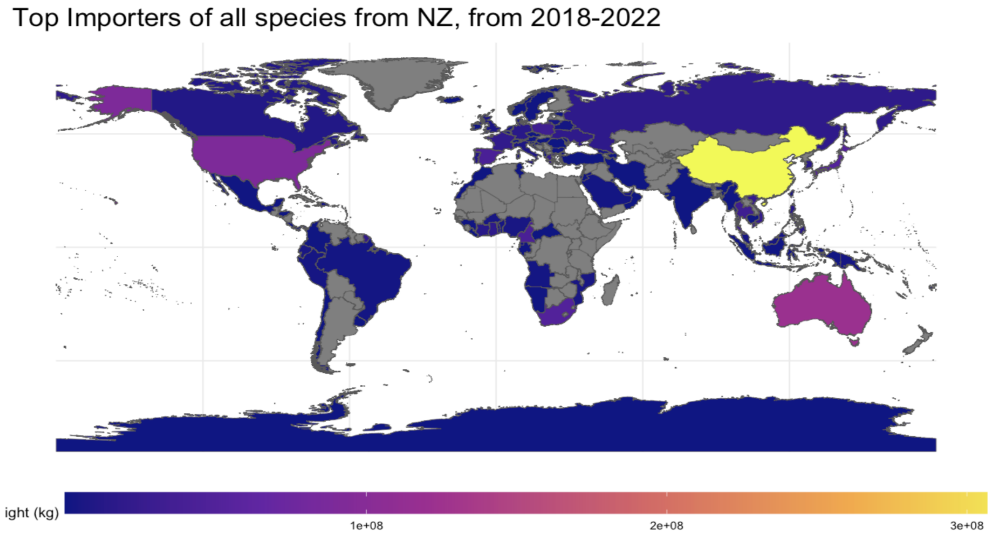

Figure 6: Top Importers of All Species

The top export markets in the world (from New Zealand) are China, Australia, USA, Japan and South Africa.

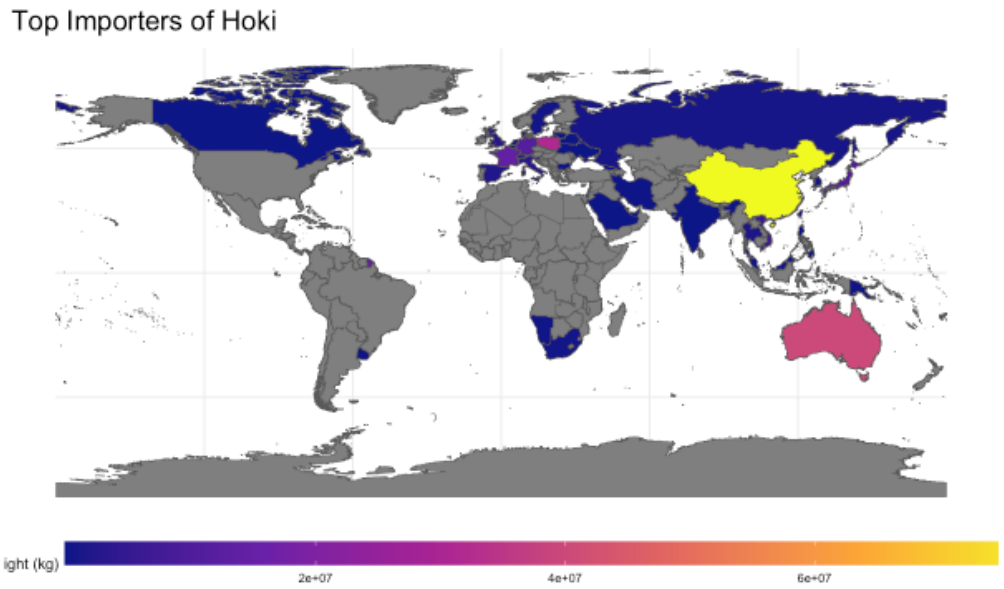

Figure 7: Top Importers of Hoki

From the chloropleth, the top importers of Hoki are China, Australia and Poland, France, Germany.

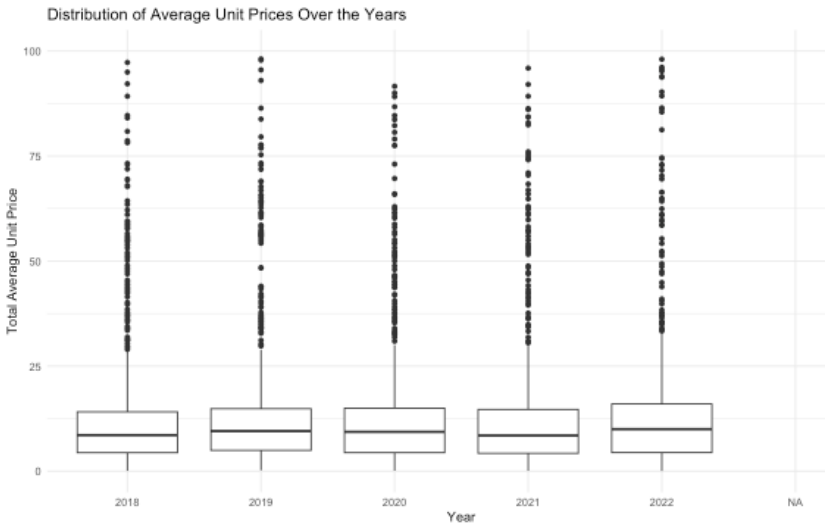

Figure 8: Distribution of Average Unit Price over the Years

Over the years, there has been almost no change in the AvgUnitPrice (of all species).

1.2 Analysis of Catch volume, Export Value, Mean SST and Fuel Prices using Julia

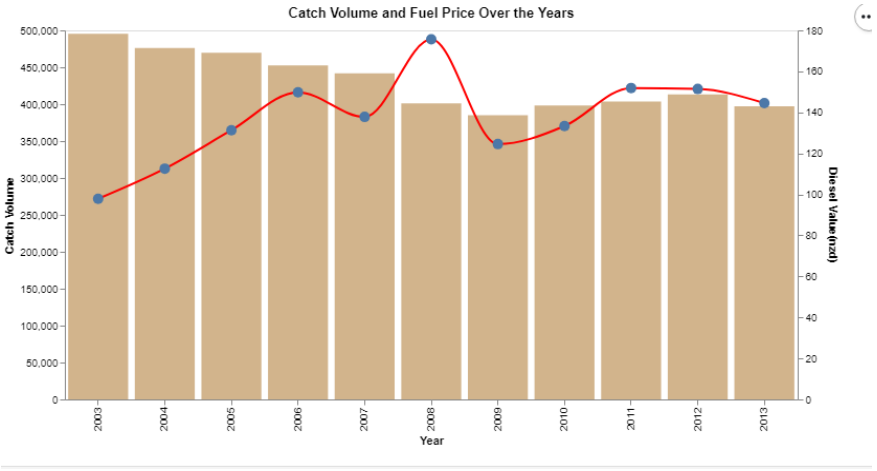

Figure 9: Catch Volume and Fuel Prices

In general, as fuel prices increase, the reported catch tends to decrease as can be seen from the plot. The decline in catch volume between 2003 and 2007 appears to coincide with the increase in fuel prices suggesting that as fuel became more expensive, fishing activities might have been curtailed leading to a reduced catch volume. In 2008, we see that fuel prices sharply increased, and there was a corresponding drop in catch volume. This potentially indicates that as fuel prices became more expensive, fishing activities may have reduced due to the increased operational costs. As prices surge, fishing operations might find it less economically viable to venture out as frequently or as far, hence leading to reduced catch volumes (Newstalk ZBBy 10 Jun and By).

From 2009 onwards, both catch volumes and fuel prices appear to vary, suggesting that other factors might have come into play in influencing catch volumes.

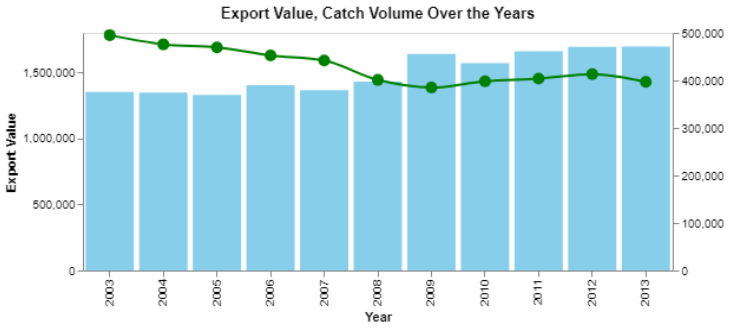

Figure 10: Export Value and Catch Volume

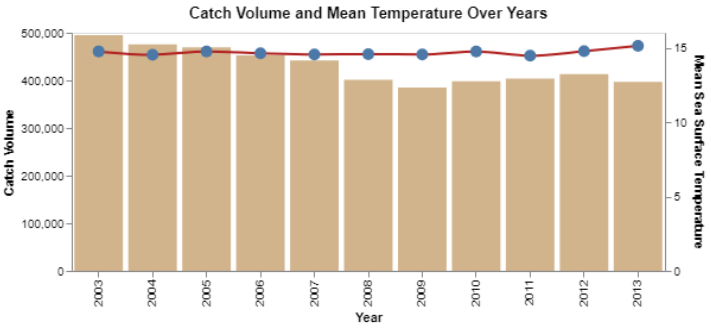

Figure 11: Catch Volume and Mean Temperature

From Figure 10, we see that catch and export value has an almost inverse relationship. As catch volume decreases, export value increases across the years. This could be due to various reasons and not just demand and supply. A lower catch volume could mean that only the best or most valuable species are being caught and exported, thus driving up the average export value. Governments might impose stricter regulations to prevent overfishing, leading to decreased catch volumes. This would need to be analysed by joining other data.

From Figure 11, the catch volume doesn't seem to change significantly with change in mean sea surface temperature. However, from 2011 the mean SST has been increasing gradually and there may be some patterns post these years. It is to be noted that the data range is spanning only 11 years and temperatures have been rising since then. Mean SST does play an important role in the spawning and growth of many marine species. Elevated sea surface temperatures can impact the distribution of species with some migrating towards cooler waters. This migration can affect the availability of certain species in traditional fishing areas.

CONCLUSION AND FUTURE SCOPE

By using appropriate data wrangling methods and techniques, we were able to investigate meaningful patterns and relationships from the New Zealand fisheries data. The significance of understanding catch volumes, pricing trends and their potential impacts on marine ecosystems were brought to the forefront. It also highlighted the fact that to conduct a meaningful analysis, the structure and quality of the data needs to be addressed. In conclusion, from the analysis of the wrangled data, the relations observed between factors like total reported catch, temperatures and price trends emphasized the economic and environmental implications of these fishing practices.

During data collection, we also scraped detailed metrics concerning the length and weight of various fish species using HTTP, Gumbo, Cascadia packages in Julia. This dataset had parameters such as minimum and maximum lengths (in centimetres) and weights (in kilograms) for 11 of the most popular New Zealand fish species, ranging from Alfonsino to Squid. This data can be joined to the main dataset using the species name. It provided insights into the physical characteristics of these marine species, however, we decided to keep it separate from our primary dataset. We wanted to maintain a more focused analysis on our main objectives. This fish metric dataset holds significant potential for future analyses. For instance, it could be used to study the correlation between environmental conditions and average fish sizes or understanding the catch patterns concerning the maturity of fish.

We also obtained information on seasonality details of New Zealand which was scraped. This data was not utilized in the current phase of our analysis, but we think it has potential value for future research. Seasonality data can offer insights into periodic fluctuations and trends in temperatures which can have profound effects on catch volumes, fish behaviours, and other related metrics. As climate change continues to have pronounced impacts on our oceans, understanding these seasonal variations becomes even more important.

CHALLENGES AND DIFFICULTIES

Inconsistencies in Naming Conventions: One of the primary challenges was dealing with inconsistent naming conventions across the datasets. For instance, the same species of fish was labelled differently in the CATCH data (Blue Cod) compared to the EXPORT data (Cod, Blue). Such differences made direct merging of datasets difficult and thus needed data cleaning.

Missing or Incomplete Data: The information we were dealing with often had gaps or missing entries. The CATCH data was missing for a lot of species. Since the CATCH data and the EXPORT

data had slightly different species, there were some fish species for which we had catch volumes, but no export information. The vice versa was also true. For example, the EXPORT data had “Herrings”, but there’s no CATCH information about that species of fish.

Integration with Taxonomy Data: The integration of taxonomy or reference data acted as a 'middle table' to bridge different naming conventions. This required a thorough understanding of marine biology classifications and a lot of manual cross-referencing.

Temporal Mismatch in Datasets: Different datasets were recorded at varying time intervals. So, obtaining datasets that correspond to the same time period was difficult. Some were annual while others were monthly. The same issues were faced in Julia datasets as the timelines of data didn’t match and so we ended up having to remove a lot of years to get valid data.

Data Source Variability: The data came from multiple sources each with its own standards and structures. This meant that even before the actual wrangling process began we had to spend a significant amount of time to understand the unique characteristics of each dataset.

In Julia, manipulating original files was a bit challenging because of merged columns and rows. Another challenge we faced was while plotting correlations using the Plot package but it was resource-intensive and we could not successfully use this package. The dataset obtained using Julia is small and may not be particularly useful for predictive analysis. The catch and export values are only available from 2003 whereas fuel prices are available from 1974. Collecting data for a wider time period may help in this case.

REFERENCES

Newstalk ZBBy 10 Jun, and By. “Fishing Businesses ‘crippled’ by Fuel Costs.” *NZ Herald*, NZ Herald, 18 Sept. 2020, www.nzherald.co.nz/business/fishing-businesses-crippled-by-fuel-costs/NRX5Z2M2E3YOEFOJNHMAMZG3WI/ .

New Zealand Ministry for Culture and Heritage Te Manatu Taonga. “Exclusive Economic Zones.” *Te Ara Encyclopedia of New Zealand – Te Ara Encyclopedia of New Zealand*, Ministry for Culture and Heritage Te Manatu Taonga, 30 Apr. 2019, teara.govt.nz/en/map/33830/exclusive-economic-zones.

Salmonidae (2022) *NIWA*. Available at: https://niwa.co.nz/freshwater/nzffd/NIWA-fish-atlas/fish-species/salmonidae

Galaxiidae (2022) *NIWA*. Available at: https://niwa.co.nz/freshwater/nzffd/NIWA-fish-atlas/fish-species/galaxiidae

Thomas, J. (30 May 2022) *Which saltwater fish are edible?*, *eHow*. Available at: https://www.ehow.com/list_7554059_saltwater-fish-edible.html.

Ministry for Primary Industries, M. for P. (10 March 2023) *Hoki: New Zealand’s largest fishery: NZ Government*, *Ministry for Primary Industries*. Available at: https://www.mpi.govt.nz/fishing-aquaculture/fisheries-management/fish-stock-status/hoki-new-zealands-largest-fishery/

To interpret the fish names in the export data :

Nature’s gift, harvested with care (2023) *Species - Seafood NZ*. Available at: https://www.seafood.co.nz/species

Mayor, D. (2023) *Sardines*, *AZ Animals*. Available at: https://a-z-animals.com/animals/sardines/

Latham, E. (22 May 2010) *All in the name of Snapper*, *Stuff*. Available at: https://www.stuff.co.nz/nelson-mail/editors-picks/3709145/All-in-the-name-of-snapper

Niwa. P. J. McMillan, M. P. Francis, G. D. James, L. J. Paul, P. Marriott, E. J. Mackay, B. A. Wood, D. W. Stevens, L. H Griggs, S. J. Baird , C. D. Roberts, A. L. Stewart, C. D. Struthers, J. E. Robbins (2019) Available at: https://docs.niwa.co.nz/library/public/NZAEBR-208.pdf

ChatGPT was used to troubleshoot some errors encountered.