

# Insights into the Fishing Industry

Erin Miller (em4919), Theresa Van (ttv2006), Vaidisha Mehta (vsm2036), Yami Naik (yn2224)



## Motivation:

Overfishing is responsible for a variety of serious downstream effects, including economic losses in related industries and environmental effects like loss of ecosystem diversity, food security, and economic output. By studying fish populations, tracking fishing vessel activities, and following economic indicators, we will glean a deeper insight into how correlated intensive fishing truly is with these various issues. Further, we hope to get a better understanding of the causes of overfishing, and to address the complex dynamics of overfishing and develop strategies to mitigate its effects.



# Overfishing: Causes & Effects

## CAUSES

- Lack of Effective Fisheries Management
- Government subsidies fueling overfishing by increasing fishing capacity
- Illegal, Unreported, and Unregulated (IUU) Fishing
- Advanced fishing technology leading to overfishing by surpassing fish reproduction rates
- Market Demand and Consumer Preferences
- Lack of Alternative Livelihoods
- Climate Change and Environmental Factors

## EFFECTS

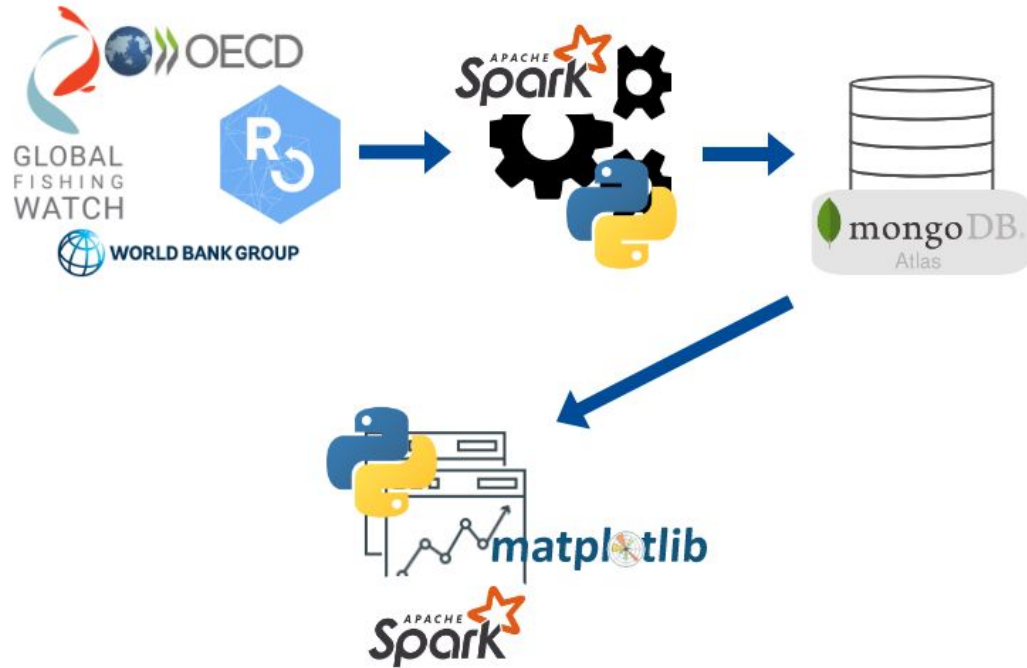
- It reduces fish populations and disrupts the ecological balance.
- Fishing vessels damage sensitive habitats, harming biodiversity.
- Fishing boats pollute and harm marine life.
- Herbivorous fish removal harms reefs, causes algae growth & coral degradation.
- Fishing lights disrupt marine behavior and migration.
- Ghost nets kill marine life, harming ecosystems.
- Fishing activity could disrupt tourism, harming the economy



## Datasets

- 1) Global Fishing Watch Events: It includes fishing events tracked by GFW, with location & vessel information; and merged based on country and month/year.
- 2) World Bank Development Indicators Tourism: Here, the tourism dataset comprises data on arrivals, receipts (in US dollars), and GDP, all of which have been merged based on country and year.
- 3) RAM Legacy Stock Assessment: Here, there are fisheries reported stats on different stocks.
- 4) OECD: It includes different nations' supply of fishing vessels and government subsidies to the fishing sector

# Technologies Used





# **ETL (Extract, Load, Transform)**





# ETL: Fishing Subsidies (General Approach)

## Issues:

- Data set up for views – a lot of repeating data
- Government support and fishing sector payments broken up into different variables (i.e. direct support, support for R&D, etc.)

## Solution:

- Aggregate subsidy variables into total support and total payments

# ETL: Fishing Subsidies (Original Dataset)

COUNTRY	Country	VARIABLE	Variable	MEASURE	Measure	YEAR	Year	Unit Code	Unit	PowerCod	PowerCod	Reference	Reference	Value
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2014	2014	USD	US Dollar	0	Units			1103477
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2015	2015	USD	US Dollar	0	Units			2045669
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2016	2016	USD	US Dollar	0	Units			1454654
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2017	2017	USD	US Dollar	0	Units			2030833
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2018	2018	USD	US Dollar	0	Units			5976.148
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2019	2019	USD	US Dollar	0	Units			0
AUS	Australia	IFINP	I.A. Transf	USD	US dollar	2020	2020	USD	US Dollar	0	Units			0
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2014	2014	USD	US Dollar	0	Units			1103477
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2015	2015	USD	US Dollar	0	Units			2045669
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2016	2016	USD	US Dollar	0	Units			854804.2
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2017	2017	USD	US Dollar	0	Units			881304.9
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2018	2018	USD	US Dollar	0	Units			2988.074
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2019	2019	USD	US Dollar	0	Units			0
AUS	Australia	IFINP_VAR	I.A.1. Tran	USD	US dollar	2020	2020	USD	US Dollar	0	Units			0
AUS	Australia	IFRPC	I.C. Transf	USD	US dollar	2010	2010	USD	US Dollar	0	Units			2353749
AUS	Australia	IFRPC	I.C. Transf	USD	US dollar	2011	2011	USD	US Dollar	0	Units			8526723
AUS	Australia	IFRPC	I.C. Transf	USD	US dollar	2012	2012	USD	US Dollar	0	Units			5578744
AUS	Australia	IFRPC	I.C. Transf	USD	US dollar	2013	2013	USD	US Dollar	0	Units			19252192
AUS	Australia	IFRPC	I.C. Transf	USD	US dollar	2014	2014	USD	US Dollar	0	Units			7481224





# ETL: Fishing Subsidies (Aggregating Sums)

Variable Code	Variable
IFINP	I.A. Transfers based on input use
IFINP_VAR	I.A.1. Transfers based on variable input use
IFINP_FIX	I.A.2. Transfers based on fixed capital formation
IFINP_FIX_V	I.A.2.1.Support to vessel construction/purchase
IFINP_FIX_M	I.A.2.2.Support to modernisation
IFINP_FIX_O	I.A.2.3.Support to other fixed costs
IFINC	I.B. Transfers based on fishers income
IFINC_INS	I.B.1. Income support
IFINC_SIF	I.B.2. Special insurance system for fishers
IFRPC	I.C. Transfers based on the reduction of productive capacity
IFMSC	I.D. Miscellaneous direct support to individuals and companies
TIFN	I.E. Tax exemptions
FTC	I.E.1. Fuel tax concessions
MPS	I.E.2. Other tax exemptions
GSACC	II.A. Access to other countries' waters
GSINF	II.B. Provision of infrastructure
GSINF_CAP	II.B.1. Capital expenditures
GSINF_ACC	II.B.2. Subsidized access to infrastructure
GSMKG	II.C. Marketing and promotion
GSCOM	II.D. Support to fishing communities
GSEDU	II.E. Education and training
GSRND	II.F. Research and development
GSMNG	II.G. Management of resources
GSMNG_EXP	II.G.1. Management expenditures
GSMNG_STK	II.G.2. Stock enhancement programs
GSMNG_ENF	II.G.3. Enforcement expenditures
GSMSC	II.H. Miscellaneous support for services to the sector
FCRRE	III.A. Payments made by the fisheries sector, for resource access rights
FCRIN	III.B. Payments made by the fisheries sector, for infrastructure access
FCRMG	III.C. Payments made by the fisheries sector, for management, research and enforcement
FCROT	III.E. Payments made by the fisheries sector, Other



## ETL: Fishing Subsidies (Data Transformation)

```
root
|-- Country Code: string (nullable = true)
|-- Country: string (nullable = true)
|-- Variable Code: string (nullable = true)
|-- Variable: string (nullable = true)
|-- Year: integer (nullable = true)
|-- Unit Code: string (nullable = true)
|-- Unit: string (nullable = true)
|-- Value: double (nullable = true)
```



```
root
|-- Country Code: string (nullable = true)
|-- Country: string (nullable = true)
|-- Year: integer (nullable = true)
|-- Unit Code: string (nullable = true)
|-- Total Support: double (nullable = true)
|-- Total Payments: double (nullable = true)
```



# ETL: Fishing Subsidies (MongoDB)

```
_id: ObjectId('645c226be08d897f174b233c')  
Country Code: "ARG"  
Country: "Argentina"  
Year: 2010  
Unit Code: "USD"  
Total Support: 12752095.02  
Total Payments: 6436043.01
```



# ETL: Fishing Fleets (General Approach)

## Issues:

- Data set up for views – a lot of repeating data
- Count of vessels broken up by vessel size

## Solution:

- Add columns for the different vessel sizes
- Nest different metrics into vessel types

# ETL: Fishing Fleets (Original Dataset)

COUNTRY	Country	FLEET	Fleet	MEASURE	Measure	YEAR	Year	Unit Code	Unit	PowerCod	PowerCod	Reference	Reference	Value
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2010	2010	NBR	Number	0	Units			318
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2011	2011	NBR	Number	0	Units			325
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2012	2012	NBR	Number	0	Units			318
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2013	2013	NBR	Number	0	Units			306
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2014	2014	NBR	Number	0	Units			309
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2015	2015	NBR	Number	0	Units			298
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2016	2016	NBR	Number	0	Units			290
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2017	2017	NBR	Number	0	Units			315
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2018	2018	NBR	Number	0	Units			285
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2019	2019	NBR	Number	0	Units			294
AUS	Australia	TOT_VESS	Total Vess	NUM	Number of	2020	2020	NBR	Number	0	Units			302
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2010	2010	TONNE	Tonnes	0	Units			36365
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2011	2011	TONNE	Tonnes	0	Units			34931
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2012	2012	TONNE	Tonnes	0	Units			35360
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2013	2013	TONNE	Tonnes	0	Units			35713
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2014	2014	TONNE	Tonnes	0	Units			40741
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2015	2015	TONNE	Tonnes	0	Units			53948
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2016	2016	TONNE	Tonnes	0	Units			64623
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2017	2017	TONNE	Tonnes	0	Units			40648
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2018	2018	TONNE	Tonnes	0	Units			39367
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2019	2019	TONNE	Tonnes	0	Units			53259
AUS	Australia	TOT_VESS	Total Vess	GRT	Gross tonn	2020	2020	TONNE	Tonnes	0	Units			59120
AUS	Australia	LOA_UNK	LOA unknc	NUM	Number of	2010	2010	NBR	Number	0	Units			14
AUS	Australia	LOA_UNK	LOA unknc	NUM	Number of	2011	2011	NBR	Number	0	Units			11



## ETL: Fishing Fleets (Expanding Vessel Sizes, resolving different metrics)

Fleet Class
0-5.9 m
12-17.9 m
18-23.9 m
24-29.9 m
30-35.9 m
36-44.9 m
45-59.9 m
6-11.9 m
60-74.9 m
75 m and over
LOA unknown
Total Vessels

Unit
Number
Tonnes



# ETL: Fishing Fleets (Data Transformation)

```
root
|-- Country Code: string (nullable = true)
|-- Country: string (nullable = true)
|-- Year: integer (nullable = true)
|-- Fleet Class: string (nullable = true)
|-- Number: integer (nullable = true)
|-- Tonnes: integer (nullable = true)
```



```
root
|-- Country Code: string (nullable = true)
|-- Country: string (nullable = true)
|-- Year: integer (nullable = true)
|-- 0-5_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 12-17_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 18-23_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 24-29_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 30-35_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 36-44_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 45-59_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 6-11_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 60-74_9 m: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- 75 m and over: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- LOA unknown: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
|-- Total Vessels: struct (nullable = true)
|   |-- Number: integer (nullable = true)
|   |-- Tonnes: integer (nullable = true)
```





# ETL: Fishing Fleets (MongoDB)

```
_id: ObjectId('645c29e90e0ea842ed1f1161')
Country Code: "ARG"
Country: "Argentina"
Year: 2010
▼ 0-5_9 m: Object
    Number: 33
▼ 12-17_9 m: Object
    Number: 287
    Tonnes: 4222
▼ 18-23_9 m: Object
    Number: 106
    Tonnes: 5396
▼ 24-29_9 m: Object
    Number: 111
```

---





# ETL: Fish Stocks (General Approach)

## Issues:

- Database set up for views
- Time series data
- Multiple tables with different foreign keys
- Very verbose database – takes up a lot of space
- Stats are not consistent across fish stocks
- Reporting years are not consistent across fish stocks
- Lots of different metrics used

## Solution:

- Create array of nested objects
- Each row will pertain to a unique fish stock ID
- Only include data from 1990 and onwards



# ETL: Fish Stock (Data Transformation)

```
root
|-- Stock ID: string (nullable = true)
|-- Common Name: string (nullable = true)
|-- Country: string (nullable = true)
|-- Country Code: string (nullable = true)
|-- Year: integer (nullable = true)
|-- Measure ID: string (nullable = true)
|-- Measure description: string (nullable = true)
|-- Unit: string (nullable = true)
|-- Value: double (nullable = true)
```



```
root
|-- Stock ID: string (nullable = true)
|-- Common Name: string (nullable = true)
|-- Country: string (nullable = true)
|-- Country Code: string (nullable = true)
|-- Years: array (nullable = false)
|   |-- element: struct (containsNull = false)
|   |   |-- Year: integer (nullable = true)
|   |   |-- Measures: array (nullable = false)
|   |       |-- element: struct (containsNull = false)
|   |           |-- Measure ID: string (nullable = true)
|   |           |-- Measure description: string (nullable = true)
|   |           |-- Unit: string (nullable = true)
|   |           |-- Value: double (nullable = true)
```



# ETL: Fish Stock (MongoDB)

```
_id: ObjectId('645c97cda812662baef4077c')
Stock ID: "ACADRED2J3K"
Common Name: "Acadian redfish"
Country: "Canada"
Country Code: "CAN"
▼ Years: Array
  ▼ 0: Object
    Year: 1990
    ▼ Measures: Array
      ▼ 0: Object
        Measure ID: "CdivMEANC-ratio"
        Measure description: "Catch divided by mean catch"
        Unit: "ratio"
        Value: 0.163295656
```

---



## ETL: Tourism

- This ETL (Extract, Transform, Load) process involves the extraction of tourism data from three different sources, transformation of the data into a consistent format, and finally loading the cleaned and merged data into a single output file.
- This ETL process provided a clean and merged dataset of tourism data for further analysis, which includes information on tourist arrivals, tourist receipts, and GDP for various countries across different years.



# ETL: Tourism (Original Dataset)

Series Name_x	Series Code_x	Country Name	Country Code	Year	Receipts	Series Name_y	Series Code_y	Arrivals	Series Name	Series Code	GDP
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Afghanistan	AFG	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	..	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	..
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Albania	ALB	1995 [YR1995]	70000000	International tourism, number of arrivals	ST.INT.ARVL	304000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	13.322333321684
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Algeria	DZA	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	520000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	3.79999478984085
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	American Samoa	ASM	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	34000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	..
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Andorra	AND	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	..	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2.75750161326249
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Angola	AGO	1995 [YR1995]	27000000	International tourism, number of arrivals	ST.INT.ARVL	9000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	15.0000000288634
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Antigua and Barbuda	ATG	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	447000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	-4.35958738239265
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Argentina	ARG	1995 [YR1995]	2550000000	International tourism, number of arrivals	ST.INT.ARVL	2289000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	-2.84520961057079
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Armenia	ARM	1995 [YR1995]	14000000	International tourism, number of arrivals	ST.INT.ARVL	12000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	6.89999841973659
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Aruba	ABW	1995 [YR1995]	554000000	International tourism, number of arrivals	ST.INT.ARVL	912000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2.54714368704694
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Australia	AUS	1995 [YR1995]	10370000000	International tourism, number of arrivals	ST.INT.ARVL	3726000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	3.88710629393142
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Austria	AUT	1995 [YR1995]	14529000000	International tourism, number of arrivals	ST.INT.ARVL	17173000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2.66798353089548
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Azerbaijan	AZE	1995 [YR1995]	87000000	International tourism, number of arrivals	ST.INT.ARVL	93000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	-11.7999991076818
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Bahamas, The	BHS	1995 [YR1995]	1356000000	International tourism, number of arrivals	ST.INT.ARVL	3239000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	4.37875209047469
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Bahrain	BHR	1995 [YR1995]	593000000	International tourism, number of arrivals	ST.INT.ARVL	2311000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	3.92999154532021
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Bangladesh	BGD	1995 [YR1995]	..	International tourism, number of arrivals	ST.INT.ARVL	156000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	5.12127789724337
International tourism, receipts (current US\$)	ST.INT.RCPT.CD	Barbados	BRB	1995 [YR1995]	628000000	International tourism, number of arrivals	ST.INT.ARVL	927000	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2.01399800028568



## ETL: Tourism (Data Transformation)

```
Series Name_x      object
Series Code_x      object
Country Name       object
Country Code       object
Year              object
Receipts           object
Series Name_y      object
Series Code_y      object
Arrivals          object
Series Name        object
Series Code        object
GDP               object
dtype: object
```



```
Country Name      object
Country Code      object
Year             object
Receipts          object
Arrivals          object
GDP              object
dtype: object
```



## ETL: Tourism (MongoDB)

```
_id: ObjectId('64591b0bf8c148fbd31e8c8f')  
Country Name: "Belarus"  
Country Code: "BLR"  
Year: "1995 [YR1995]"  
Receipts: "280000000"  
Arrivals: ".."  
GDP: "-10.4000004914994"
```



# ETL: Global Fishing Watch (General Approach)

## Issues:

- Obscure MRGID EEZ codes
- Poor documentation for how to query API
- Lining up with other datasets

## Solution:

- Wrote R script using mregions package to get EEZ codes
- Lots of trial and error
- First tried using coordinates grid, then used EEZ codes





# ETL: Global Fishing Watch (Mongo)

```
_id: ObjectId('645b9acd59f1d22bb91c872c')
year: 2013
month: "01"
country: "Belgian Exclusive Economic Zone"
▼ record: Object
  id: "3836be206e345835cf838ce9efddb9bf"
  type: "fishing"
  start: "2013-01-02T00:12:49.000Z"
  end: "2013-01-02T06:58:29.000Z"
▼ position: Object
  lat: 51.6458
  lon: 2.857
▼ regions: Object
  ▼ mpa: Array
  ▼ eez: Array
    0: "3293"
    1: "5668"
  ▼ rfmo: Array
    0: "IWC"
    1: "NEAFC"
    2: "NAMMCO"
    3: "ICES"
    4: "ACAP"
    5: "ICCAT"
    6: "NASCO"
  ▼ fao: Array
    0: "27.4"
    1: "27"
    2: "27.4.c"
  ▼ majorFao: Array
    0: "27"
  ▼ eez12Nm: Array
  ▼ highSeas: Array
  ▼ mpaNoTakePartial: Array
  ▼ mpaNoTake: Array
▼ boundingBox: Array
  0: 2.9316116667000074
  1: 51.5670983333
  2: 2.8046133333000114
  3: 51.7502766667
▼ distances: Object
  startDistanceFromShoreKm: 30
  endDistanceFromShoreKm: 49
  startDistanceFromPortKm: 21.755406
  endDistanceFromPortKm: 18.835493999999997
▼ vessel: Object
  id: "9262f2c51-11ac-6ab9-b7a1-cf5d5d3c6cb6"
  flag: "NLD"
  name: "FV ARM20"
  ssvId: "245473000"
  ► authorizations: Array
▼ fishing: Object
  totalDistanceKm: 48.42475193625251
  averageSpeedKnots: 5.063999981884001
  averageDurationHours: 0.28171296296296294
  potentialRisk: false
  vesselAuthorizationStatus: "unauthorized"
```



# Data Analysis

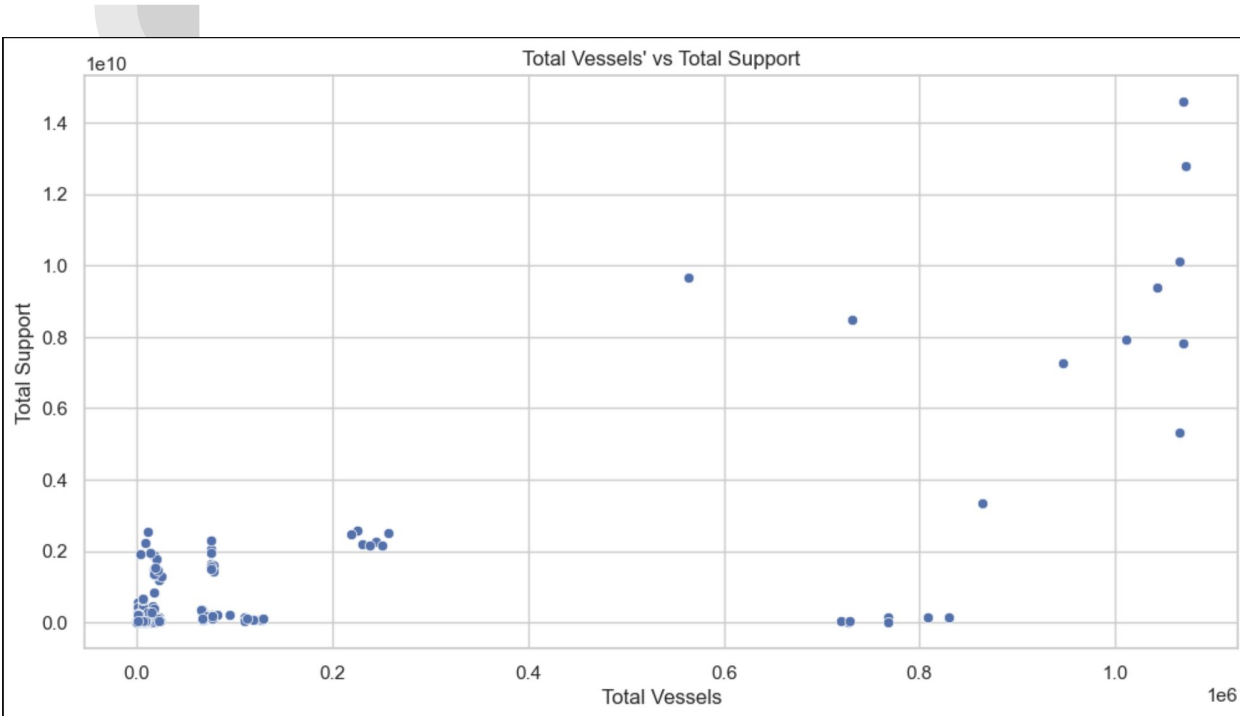




## Use Cases: Trend Between Vessels and Subsidies:

```
vessel_subsidies_df.dtypes
```

```
[('Country', 'string'),  
 ('Country Code', 'string'),  
 ('Year', 'int'),  
 ('0-5_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('12-17_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('18-23_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('24-29_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('30-35_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('36-44_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('45-59_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('6-11_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('60-74_9 m', 'struct<Number:int,Tonnes:int>'),  
 ('75 m and over', 'struct<Number:int,Tonnes:int>'),  
 ('LOA unknown', 'struct<Number:int,Tonnes:int>'),  
 ('Total Vessels', 'struct<Number:int,Tonnes:int>'),  
 ('_id', 'struct<oid:string>'),  
 ('Total Payments', 'double'),  
 ('Total Support', 'double'),  
 ('Unit Code', 'string'),  
 ('_id', 'struct<oid:string>')]
```



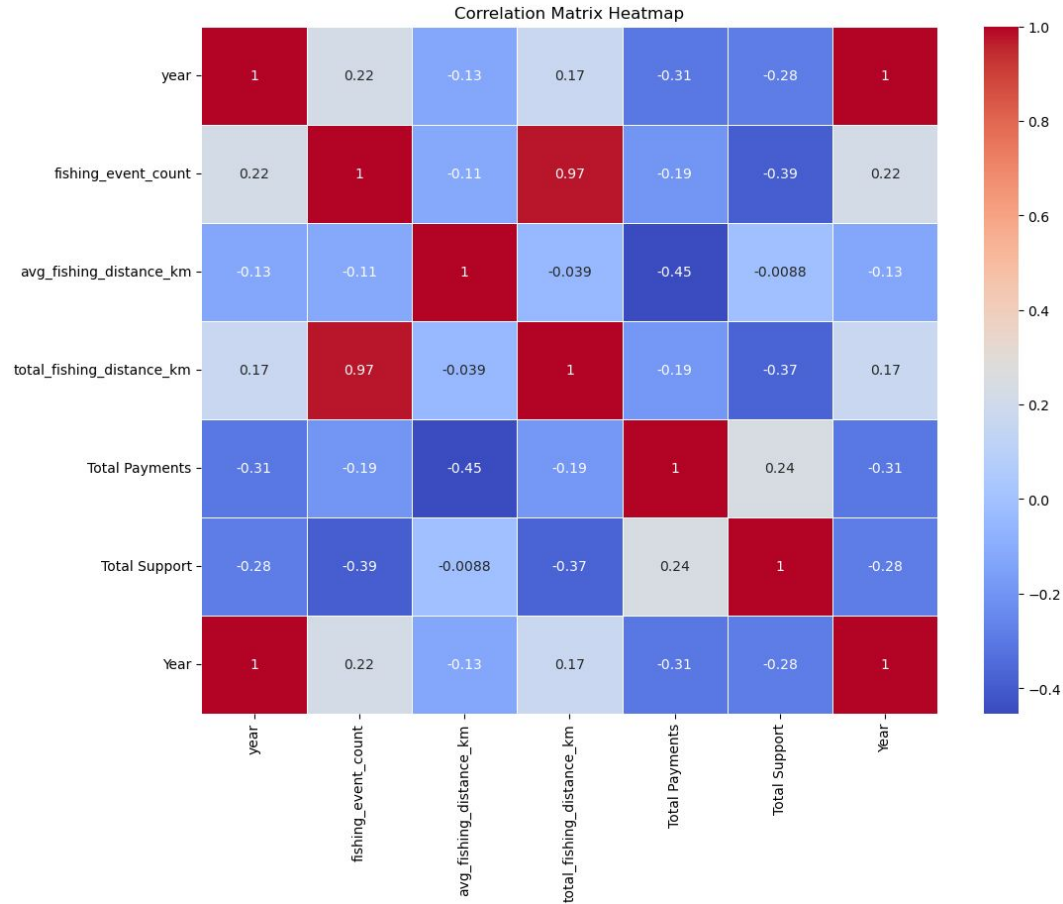
- In the analysis, we found out that the pearson correlation coefficient is equivalent to 0.7585 which shows a strong relationship vessels and support.
- This means that as one variable increases, the other variable tends to increase as well.



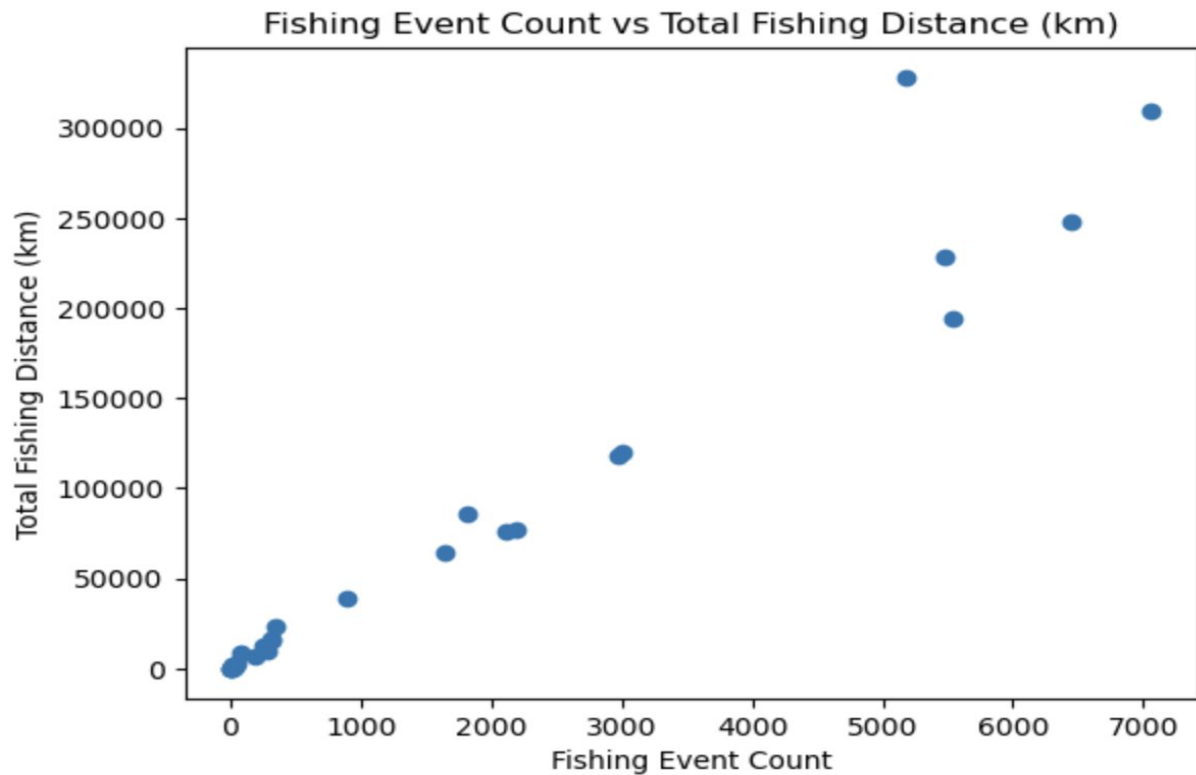
## Trend Between Subsidies and Global Fishing Watch

```
subsidies_gfw_df.dtypes
```

```
[('flag', 'string'),  
 ('year', 'int'),  
 ('fishing_event_count', 'bigint'),  
 ('avg_fishing_distance_km', 'double'),  
 ('total_fishing_distance_km', 'double'),  
 ('Country', 'string'),  
 ('Country Code', 'string'),  
 ('Total Payments', 'double'),  
 ('Total Support', 'double'),  
 ('Unit Code', 'string'),  
 ('Year', 'int'),  
 ('_id', 'struct<oid:string>')]
```



We investigated the correlation between Total Payments and Fishing Event Count, which was  $-0.19282277939665396$ , and between Total Support and Fishing Event Count, which was  $-0.39494837337751765$ . Both relationships were negative. To further visualize these connections, we incorporated a heatmap. This revealed a significant relationship between Fishing Event Count and Total Fishing Distance.



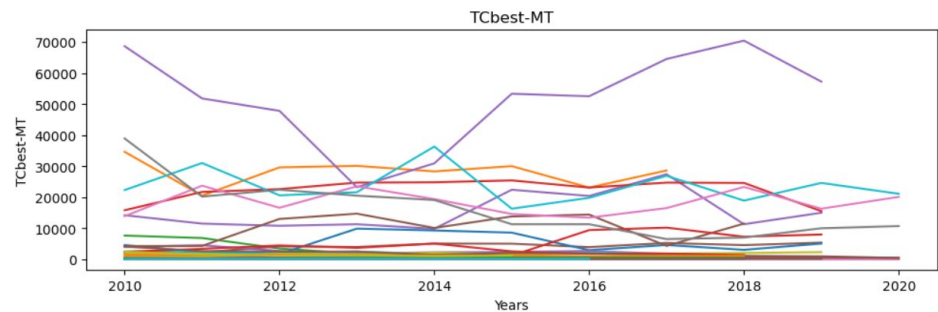
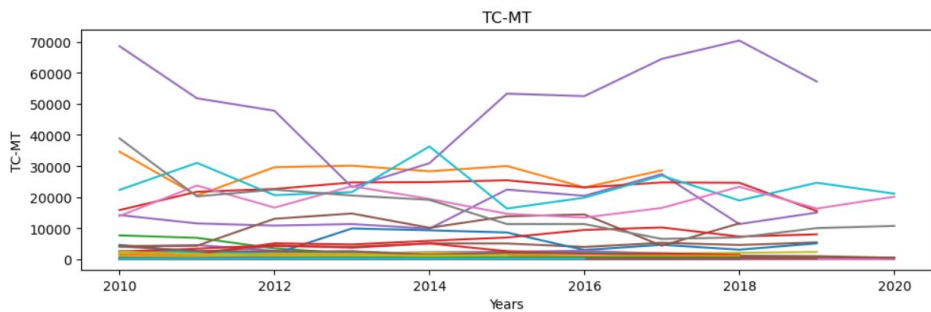
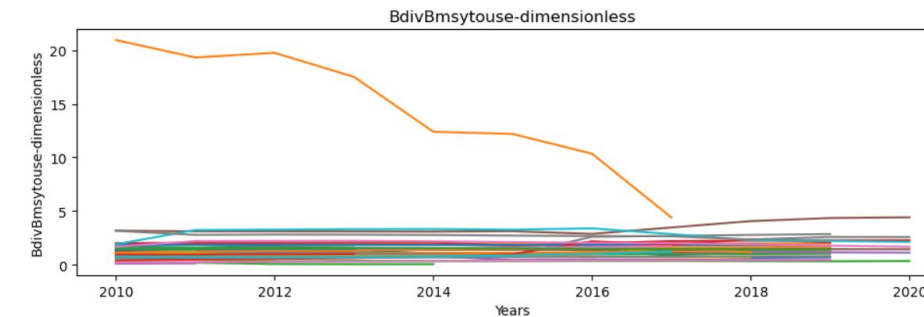
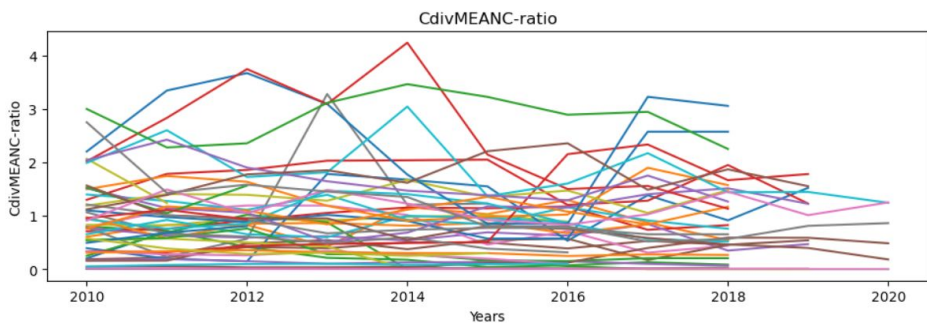
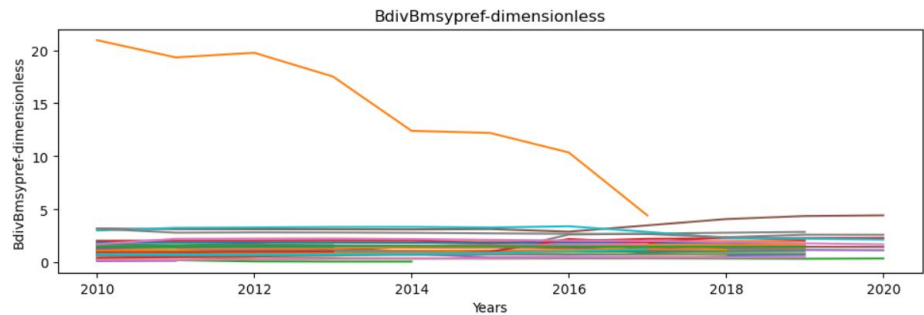
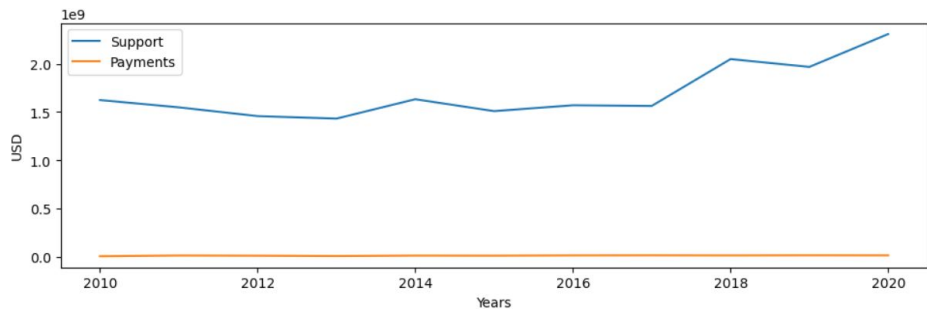
The graph shows relationship between fishing event count and total fishing distance.



# Use Case: Trends between Subsidies and Fish Stock Stats

- Visualize a general overview of government subsidies and fish stock stats
- Used sample of 50 unique fish stocks in the US with dates ranging between 2010-2020
  - Originally pulled 100 US fish stocks but after querying for only stats between 2010-2020 only 50 remained
- Measures analyzed:
  - BdivBmsytouse-dimensionless (General biomass time series relative to mean square yield reference point)
  - BdivBmsypref-dimensionless (General biomass time series preferentially relative to MSY-BRP, otherwise management target)
  - CdivMEANC-ratio (Catch divided by mean catch)
  - TCbest-M (General total Catch)
  - TC-MT (Total catch, i.e., landings and discard)

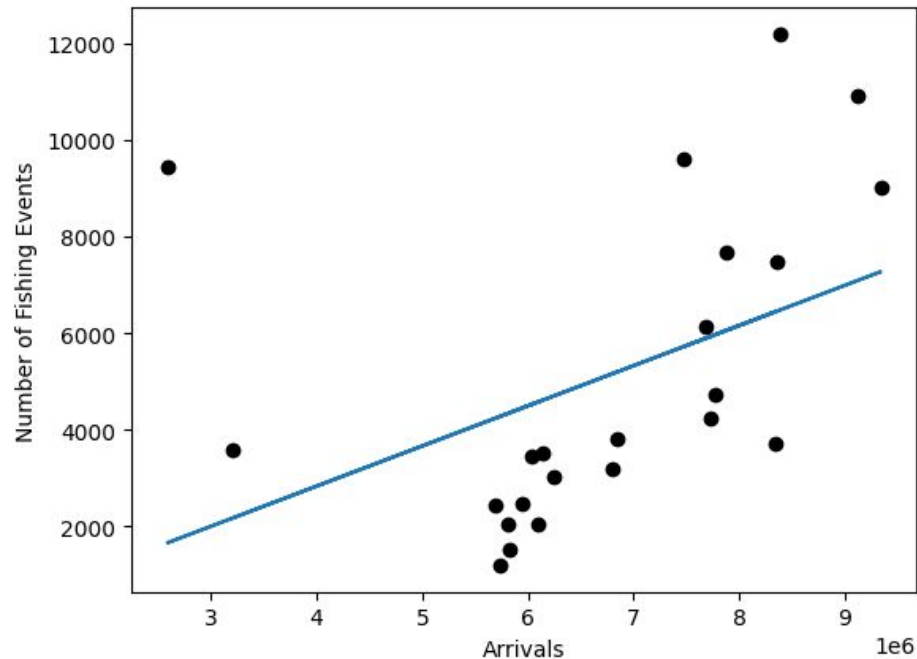






# Use Case: Trend Between GFW Events & Tourism

Fishing Events Correlation to Arrivals: 0.42463747529299817





## Future Scope:

- Do further analysis by using ocean surface temperature dataset and light pollution caused by fishing boat dataset.
  - Issues with data availability and formats
- Analyze global fishing watch data from a greater number of countries
- Find fishing subsidies dataset for non-OECD countries



**Thank you!**