AIS Data: Most-Visited Marinas

Author: Water Revolution Foundation | Sub-Study to YETI JIP | 23 October, 2024

This research has been conducted in support of the development of the Yacht Environmental Transparency Index (YETI) by a Joint Industry Project (JIP) organised through Water Revolution Foundation. YETI is the result of an extensive collaboration between renowned shipyards, naval architecture studios and research institutes, and offers a systemic, data-driven approach to assess and compare the environmental performance of large yachts A shared motivation to educate on and visualize impacts and solutions has driven their efforts to develop this tool and formed the basis for the research presented below…

---

1. Abstract

For this research, AIS data acquired in 2020 by Water Revolution Foundation was utilized to identify the world's most frequently visited marinas. By analyzing the frequency of these visits, we can weigh the shore power environmental impact per country according to the frequency of marina visits. This approach complements the calculations made by the TETIS Institute, which are based on each country's land grid, by providing a more nuanced assessment.

2. Location

This data is comprised of the following variables:

- MMSI
- IMO
- SHIPNAME
- TYPE
- DRAUGHT
- Date
- Time
- EVENT
- PORTNAME
- LOCODE
- LENGTH
- WIDTH

The dataset under analysis is extensive, consisting of 143,101 rows of data that provide a comprehensive overview of maritime activity over a significant period. Within this dataset, there are 1,539 unique ship names, each corresponding to a vessel tracked over time. These ship names are associated with 1,406 unique IMO numbers, indicating a broad range of vessels, some of which may have undergone name changes or other modifications over the data collection period. In any case, these were filtered and removed, ending up with a sample of 1082 unique vessels.

The temporal coverage of the dataset is substantial, beginning on December 24, 2017, at 00:25:15, and extending through to December 31, 2020, at 23:06:39. On average, each ship name appears

approximately 92.98 times in the dataset, which suggests regular tracking and updates for the majority of vessels. The average timespan between consecutive entries is 7 days and 12 hours.
In terms of timeframes, the data indicates that the average duration of entries for each unique ship name is about 675.74 days. This highlights the long-term monitoring of vessels, allowing for an in-depth examination of their movements and activities over nearly two years per vessel, on average.

Regarding the physical characteristics of the vessels, the average length of ships documented in the dataset is approximately 48.32 meters. In figure 1, the distribution of the fleet per size in meters is given.

Furthermore, the dataset is geographically expansive, encompassing data from 1,349 unique ports, covering countries in the Mediterranean, Caribbean, North America and Asia. This wide range of locations illustrates the global reach and operational scope of the vessels tracked, covering a diverse set of maritime hubs and destinations.

An important limitation of this dataset is the absence of information regarding the vessels' flags, which makes it challenging to determine the criteria used for selecting these particular vessels. As a result, it is difficult to assess whether this dataset accurately represents a comprehensive sample of the global fleet.

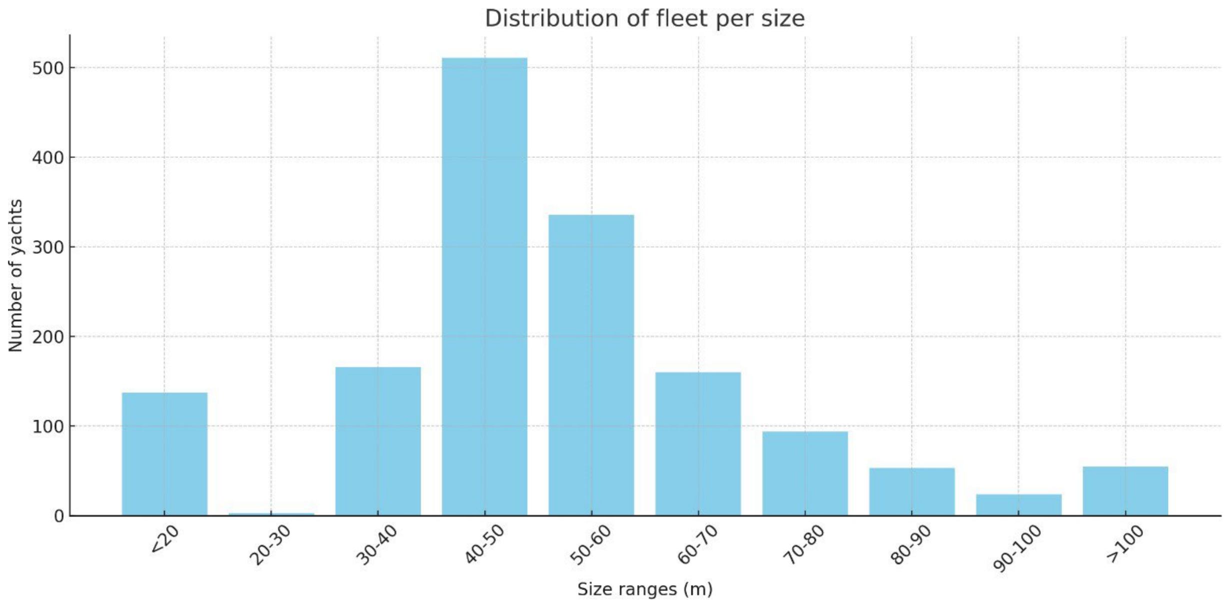


Figure 1: Distribution of fleet per size in length (m)

2.1 Data Sample

SHIPNAME	day	month	year	time	EVENT	PORTNAME	LOCODE
Example	18	4	2019	11:31:38	ENTER_ZONE	Gibraltar	GIGIB
Example	20	4	2019	10:24:09	EXIT_ZONE	Gibraltar	GIGIB
Example	20	4	2019	10:24:17	ENTER_ZONE	Gibraltar	GIGIB
Example	20	4	2019	13:43:39	EXIT_ZONE	Gibraltar	GIGIB
Example	23	7	2019	15:05:12	ENTER_ZONE	Porto-Vecchio	FRPVO
Example	24	7	2019	07:23:47	EXIT_ZONE	Porto-Vecchio	FRPVO
Example	26	7	2019	15:06:25	ENTER_ZONE	Porto-Vecchio	FRPVO

Example	29	7	2019	11:38:06	EXIT_ZONE	Porto-Vecchio	FRPVO
Example	9	8	2019	10:31:34	ENTER_ZONE	Cannes	FRCEQ
Example	20	8	2019	10:43:46	EXIT_ZONE	Cannes	FRCEQ
Example	20	8	2019	12:56:16	ENTER_ZONE	Nice	FRNCE
Example	22	8	2019	06:18:31	EXIT_ZONE	Nice	FRNCE
Example	30	8	2019	14:31:31	ENTER_ZONE	Nice	FRNCE
Example	31	8	2019	07:27:55	EXIT_ZONE	Nice	FRNCE
Example	31	8	2019	14:49:33	ENTER_ZONE	Nice	FRNCE
Example	1	9	2019	10:11:34	EXIT_ZONE	Nice	FRNCE

In the table above, it is possible to deduct that reading entries only from the set of data will not clearly indicate the correct number of visits of a certain marina. For the port ofGibraltar, Porto – Vecchio and Nice, there are multiple entries and exits even though it is most probable that the vessel never left the marina. For this reason, the possibility of searching the data for unique entries was deemed unsuitable.

3. Logic Implementation

The dataset for AIS tracking was analyzed using Python, and due to its complexity, a structured approach was necessary to efficiently count and track the number of uniquevisits within a specified timeframe. To achieve this, a custom logic was implemented to organize the data and visualize the total number of visits per location.

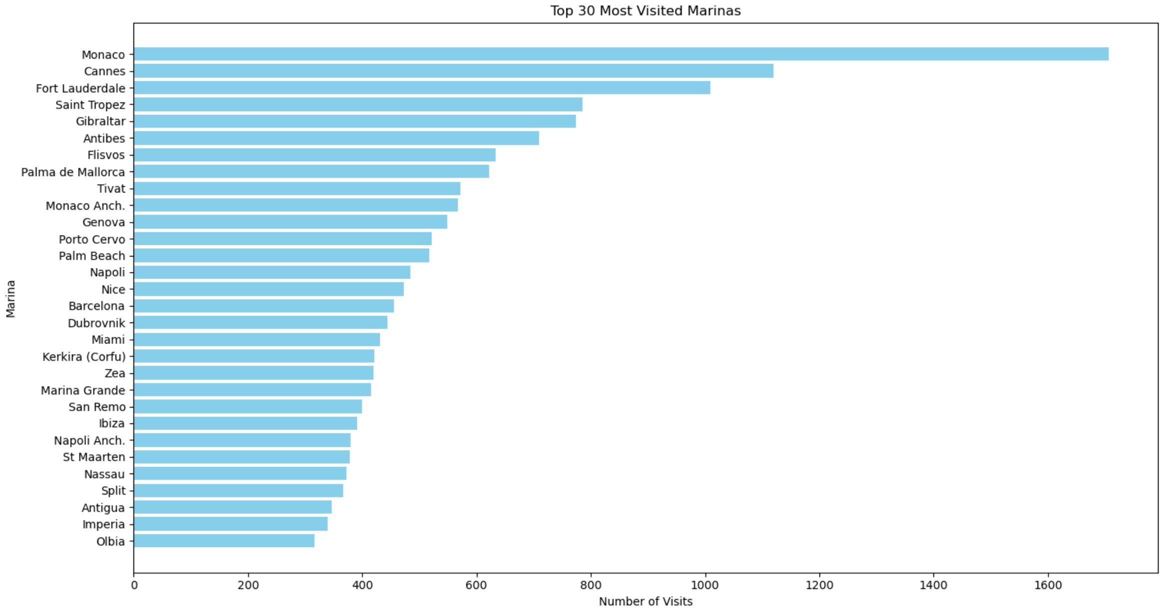
The process follows a sequential analysis of the dataset, identifying the first occurrence of a visit to each location, followed by tracking subsequent returns to thesame place as additional visits. An illustrative example is shown below:

1. Mallorca (1st visit)
2. Mallorca
3. Algiers (1st visit)
4. Algiers
5. Algiers
6. Mallorca (2nd visit)
7. Algiers (2nd visit)
8. Mallorca (3rd visit)
9. Monaco (1st visit)
10. Monaco
11. Monaco
12. Monaco
13. Antibes (1st visit)

In this sequence, visits are recorded only when a vessel revisits the same location after having visited another. Based on this, the final results are:

- Mallorca: 3 visits
- Algiers: 2 visits
- Monaco: 1 visit
- Antibes: 1 visit

Following this logic, the top 30 most visited marinas of this dataset are given in the image and table below:



	Marina	Number of Visits		Marina	Number of Visits
1	Monaco	1707	16	Barcelona	456
2	Cannes	1120	17	Dubrovnik	444
3	Fort Lauderdale	1009	18	Miami	431
4	Saint Tropez	786	19	Kerkira (Corfu)	421
5	Gibraltar	774	20	Zea	419
6	Antibes	710	21	Marina Grande	415
7	Flisvos	634	22	San Remo	400
8	Palma de Mallorca	622	23	Ibiza	391
9	Tivat	572	24	Napoli Anch.	379
10	Monaco Anch.	568	25	St Maarten	378
11	Genova	549	26	Nassau	372
12	Porto Cervo	521	27	Split	366
13	Palm Beach	517	28	Antigua	347
14	Napoli	484	29	Imperia	339
15	Nice	473	30	Olbia	317

Besides these 30 ports, the top 180 ports were taken into consideration. The results table containing the 180 ports is included in appendix A. Each of these port was attributed to a country according to its location and the list can be seen in the following page with the total number of visits per country, for a total of 37 countries.

Row Labels	Sum of Number of Visits
Abu Dhabi	52
Albania	253
Antigua and Barbuda	399
Australia	38
Bahamas	37
Canada	130
Croatia	2078
Cyprus	54
Dubai	153
Ecuador	38
Egypt	215
France	4477
French Polynesia	38
Gibraltar	839
Greece	4681
Italy	7030
Maldives	81
Malta	297
Mexico	150
Monaco	2275
Montenegro	767
Nassau	372
Netherlands	37
Norway	67
Panama	221
Portugal	55
Puerto Rico?	48
Russia	48
Saint Barthélemy	206
Singapore	76
Sint Maarten	429
Spain	2100
St Vincent and the Grenadines	222
Tunisia	53
Turkey	664
US Virgin Islands	196
USA	3186
Grand Total	32062

The table below outlines the most visited countries by yachts in this dataset, along with the number of visits and the percentage each country represents of the total. The TETIS Institute previously conducted an analysis of shore power impact per country using a lifecycle assessment (LCA) approach. While their analysis provided a solid starting point, the list of countries was somewhat limited, and the calculations were based on a simple average that didn't account for the varying frequency of yacht visits.

With this updated dataset, not only can the list of countries be expanded, but the frequency of visits can now be factored into a weighted average. This allows for a more precise evaluation of the environmental impact of shore power, reflecting the real-world patterns of yacht visits to different regions. By integrating visitation frequency into the analysis, this approach ensures a more representative and accurate calculation of shore power's overall impact across the most visited yachting destinations.

Country	**Number of visits**	
Italy	7030	21%
Greece	4681	14%
France	4477	14%
USA	3186	10%
Monaco	2275	7%
Spain	2100	6%
Croatia	2078	6%
Dominican Republic	875	3%
Gibraltar	839	3%
Montenegro	767	2%
Turkey	664	2%
Cuba	583	2%
Curacao	583	2%
Sint Maarten	429	1%
Antigua and Barbuda	399	1%
Nassau	372	1%
Malta	297	1%
Albania	253	1%
St Vincent and the Grenadines	222	1%
Panama	221	1%
Egypt	215	1%
Saint Barthélemy	206	1%
US Virgin Islands	196	1%

Geographical framework	Country	Country		Country Acronym	UF: 1 MWh	
		YES	NO		GWP [kgCO2eq/MWh]	Ecopoints [Pt/MWh]
MEDIT	Albania	1		{AL}	145,32	9,15
	Croatia	1		{HR}	451,26	27,80
	France	1		{FR}	89,78	5,28
	Gibraltar	1		{GI}	953,85	33,83
	Greece	1		{GR}	684,81	36,73
	Israel	1		{IL}	765,71	28,04
	Italy	1		{IT}	393,10	13,41
	Malta	1		{MT}	462,73	13,01
	Montenegro	1		{ME}	773,35	13,01
	Morocco	1		{MA}	1054,73	50,48
	Principality of Monaco	1		{FR}	89,78	5,28
	Spain	1		{ES}	284,55	12,03
	Tunisia	1		{TN}	758,25	17,30
	Turkey	1		{TR}	622,14	72,72
	Subtotal	14	0	Average	537,81	24,15
	Subtotal percentage	100%	0%			
CARIB	Antiqua		1	-	-	-
	Aruba		1	-	-	-
	Barbados		1	-	-	-
	Bermuda		1	-	-	-
	British Virgin Islands (BVI)		1	-	-	-
	Colombia	1		{CO}	272,55	12,85
	Cuba	1		{CU}	1208,65	46,28
	Curaçao	1		{CW}	857,89	34,79
	Dominican Republic	1		{DO}	978,88	37,13
	Grenada		1	-	-	-
	Guadeloupe - Overseas department and region of France		1	-	-	-
	Jamaica	1		{JM}	1001,67	35,89
	Panama	1		{PA}	381,04	14,56
	Puerto Rico		1	-	-	-
	Saint Lucia		1	-	-	-
	Saint Martin		1	-	-	-
	Sint Maarten - Kingdom of the Netherlands		1	-	-	-
	The Bahamas		1	-	-	-
	Trinidad and Tobago	1		{TT}	716,69	16,36
	USA - Florida	1		{SERC}	529,63	19,70
	United States Virgin Islands		1	-	-	-
	Subtotal	8	13	Average	743,37	27,20
	Subtotal percentage	38%	62%			
Overall average					612,56	25,26

With the table above, it is possible to compare which countries are included in both the new AIS database and the TETIS database.

Countries Present in Both the AIS and TETIS Databases:

- Italy
- Greece
- France
- USA
- Monaco
- Spain
- Croatia
- Dominican Republic
- Gibraltar
- Montenegro
- Turkey
- Curacao
- Sint Maarten
- Malta
- Albania
- Panama
- US Virgin Islands

Countries Present in the New AIS Database but Missing in the TETIS Database:

- Cuba
- Nassau
- St Vincent and the Grenadines
- Egypt
- Saint Barthélemy

Countries Present in the TETIS Database but Missing in the New AIS Database:

- Israel
- Morocco
- Tunisia
- Aruba
- Bermuda
- British Virgin Islands
- Colombia
- Grenada
- Guadeloupe
- Jamaica
- Puerto Rico
- Saint Lucia
- Bahamas
- Trinidad and Tobago

From here, it is essential for Water Revolution and TETIS to collaborate on how best to merge these two datasets and create a comprehensive final list. This would involve researching the shore power environmental impact for the newly added countries and ultimately calculating a weighted average for the entire list based on yacht visitation patterns.

Appendix

180 most visited marinas

Marina	Country	Number of Visits
Monaco	Monaco	1707
Cannes	France	1120
Fort Lauderdale	USA	1009
Saint Tropez	France	786
Gibraltar	Gibraltar	774
Antibes	France	710
Flisvos	Greece	634
Palma de Mallorca	Spain	622
Tivat	Montenegro	572
Monaco Anch.	Monaco	568
Genova	Italy	549
Porto Cervo	Italy	521
Palm Beach	USA	517
Napoli	Italy	484
Nice	France	473
Barcelona	Spain	456
Dubrovnik	Croatia	444
Miami	USA	431
Kerkira (Corfu)	Greece	421
Zea	Greece	419
Marina Grande	Italy	415
San Remo	Italy	400
Ibiza	Spain	391
Napoli Anch.	Italy	379
St Maarten	Sint Maarten	378
Nassau	Nassau	372
Split	Croatia	366
Antigua	Antigua and Barbuda	347
Imperia	Italy	339
Olbia	Italy	317
Golfe-Juan	France	294
Mykonos	Greece	280
Santa Maria La Carita	Italy	278
Bonifacio	France	277
Spetsai	Greece	272
Newport RI	USA	258
Viareggio	Italy	256
Taormina Anch.	Italy	254
Valletta	Malta	238
La Spezia	Italy	228
Paxi	Greece	213
Gustavia	Saint Barthélemy	206
Cavtat	Croatia	202
Saint Thomas	US Virgin Islands	196
Calvi	France	186
Agios Kosmas	Greece	183

Trogir	Croatia	181
Riposto	Italy	178
Fethiye	Turkey	177
Vouliagmeni	Greece	176
Milos (Adamas)	Greece	175
Korcula	Croatia	173
Siracusa	Italy	171
Portofino	Italy	167
La Ciotat	France	166
Castellammare di Stabia	Italy	162
Poros Kefallinias	Greece	155
Hvar	Croatia	154
Perama	Greece	153
Loano	Italy	153
Livorno	Italy	151
Lacco Ameno	Italy	150
Gocek	Turkey	148
Thira	Greece	147
Idhra (Hydra)	Greece	142
Venezia	Italy	141
Rhodes	Greece	140
Panama Canal Anch. Pacific	Panama	131
Meganisi	Greece	131
As Suways (Suez) Anch.	Egypt	124
Nantucket	USA	124
Faliro	Greece	122
Portoferraio	Italy	118
Ortakent	Turkey	114
Zadar	Croatia	114
Mustique Island Anch.	St Vincent and the Grenadines	112
Argostolion	Greece	111
Bequia Anch.	St Vincent and the Grenadines	110
Lipari	Italy	108
Palermo	Italy	105
Portals Nous	Spain	104
Saint-Mandrier-sur-Mer	France	101
Port Grimaud	France	101
Bar	Montenegro	100
Rovinj	Croatia	100
Santa Cruz de Tenerife	Spain	100
Marmaris	Turkey	100
San Diego	USA	99
New York	USA	99
Siracusa Anch.	Italy	97
Mahon, Menorca	Spain	97
Fiumicino	Italy	96
Sarande	Albania	95
Pula	Croatia	95
Kotor	Montenegro	95
Civitavecchia	Italy	92
Port Said Anch.	Egypt	91
Freeport	USA	90
Syros	Greece	90

Bodrum	Turkey	89
Sorrento	Italy	87
Marseille	France	87
Jebel Ali	Dubai	86
Vancouver	Canada	85
Tarragona	Spain	83
Ajaccio	France	82
Male	Maldives	81
Key West	USA	80
Durres	Albania	79
Papeete	Albania	79
Zakynthos	Greece	77
Singapore	Singapore	76
Sibenik	Croatia	76
Kea	Greece	74
Ischia	Italy	71
Porto Heli	Greece	69
Rijeka	Croatia	68
Seriphos	Greece	68
Bergen	Norway	67
Dubai / Port Rashid	Dubai	67
Sag Harbor	USA	67
Alimos	Greece	65
Valencia	Spain	65
Gibraltar East Anch.	Gibraltar	65
Nafplion	Greece	64
Navpaktos	Greece	62
Philipsburg	USA	62
Fiskardo	Greece	61
Seattle	USA	61
Savannah	USA	61
St Georges	Malta	59
Vilanova	Spain	58
Charleston	USA	58
Marina di Carrara	Italy	57
La Paz	Mexico	57
Boston	USA	57
Cabo San Lucas	Mexico	56
Skradin	Croatia	56
Horta	Portugal	55
Genova Anch.	Italy	55
Limassol	Cyprus	54
Porto-Vecchio	Italy	54
Cagliari	Italy	53
Bizerte	Tunisia	53
Portovenere	Italy	53
La Seyne-sur-Mer	France	52
Barbuda	Antigua and Barbuda	52
Mina Zayed	Abu Dhabi	52
Marigot	Sint Maarten	51
Limon Bay Marina Anch.	Panama	50
Vibo Valentia	Italy	49
Mali Losinj	Croatia	49
Monemvasia	Greece	48

San Juan	Puerto Rico?	48
Sochi	Russia	48
Marina di Pisa	Italy	47
Ios	Greece	47
Victoria	Canada	45
Sami	Greece	43
Cartagena	Spain	43
Porto Santo Stefano	Italy	42
Beaulieu	France	42
Porto Azzurro	Italy	41
Las Palmas	Spain	41
Malaga	Spain	40
Limon Bay Marina	Panama	40
Kos	Greece	39
Portland, Maine	USA	39
Baquerizo Moreno Anch.	Ecuador	38
Bora Bora	French Polynesia	38
Brindisi	Italy	38
Cairns	Australia	38
Savona	Italy	37
Amsterdam	Netherlands	37
Chub Cay	Bahamas	37
La Paz Anch.	Mexico	37
Norfolk	USA	37
Thunderbolt	USA	37
Livorno Anch.	Italy	37
Kusadasi	Turkey	36

Deleted yachts

Deleted yachts:

IMO	Vessel Names
1000617	['THE GOOSE' 'ATLANTIC GOOSE']
1001178	['MINDERELLA' 'AMARA']
1001960	['ILCIGNO' 'IL CIGNO']
1002093	['TRANQUILLITY' 'TRANQUILITY']
1002225	['SCOUT II' 'SCOUT' 'SAM4V']
1002249	['TANIT' 'MY BROTHER BRANCUSI']
1002706	['MMM' 'KARIMA']
1002902	['MYLIN IV' 'M/Y MYLIN IV']
1003176	['M.Y.CONSTANCE' 'M.Y.CONSTA&C% X']
1003217	['STRANGELOVE' 'NOSTALGIA' 'N0STALGIA']
1003267	['ILLUSION I' 'ILLUSION']
1004675	['ATHINA III' 'ASTARTE']
1004704	['MY ALWAELI' 'ALWAELI']
1004819	['MY LADY' 'LADY M II' 'LADY M I I']

1004936	['RASSELAS' 'BROADWATER']
1005409	['Friendship' 'SUNRISE']
1005411	['LADY ELLENII' 'AZUL V']
1005435	['GITANA' 'CAROLINA']
1005746	['TANIA T' 'LADY G II']
1005837	['CORINTHIAN' 'BB4P']
1005904	['YASMINE OF THE SEA' 'MY STARGATE']
1006207	['ENDEAVOUR T/T ARCTIC' 'ARCTIC P']
1006219	['TELEOST' 'TELEOP$']
1006386	['M/Y KISSES' 'KISSES']
1006556	['PICNIC' 'BROADWATER']
1006673	['SENSES' 'CATALYST']
1006738	['V6' 'MY V6' 'MY T6']
1006855	['TRUE NORTH' 'PRAXIS']
1006867	['TUGATSU' 'MP5']
1006881	['HAMPSHIRE I' 'HAMPSHIRE' 'AMBITION']
1006893	['MARY A.' 'MARY A']
1007158	['MERCURY' 'MALIBU']
1007251	['LADY INDIA' 'LADY CHARLOTTE' 'LADY CHARLOTTE']
1007304	['CHECKERS' 'ADYTUM']
1007419	['FULL MOON' 'ANGIAMO']
1007562	['CAPRI I' 'CAPRI']
1007586	['MY FALCON' 'FALCON']
1007689	['HAPPY DOLPHIN II' 'HAPPY DOLPHIN 2']
1007782	['M/Y ANDREA' 'ANDREA']
1007940	['NITA K II' 'NITA K 2']
1007964	['MY ILONA' 'ILONA']
1007976	['VOLPINI' 'VIBRANCE' 'VIBRA']
1008035	['ZEEFAKKEL' 'A2']
1008047	['BRAVADO' 'BRAVA']
1008164	['MY WAY V' 'MY WAY']
1008190	['Samar' 'SAMAX' 'SAMAR']
1008217	['LADY S' 'LADY E']
1008243	['FORTUNATE SUN' 'FORTU' 'FORTNATE SUN']
1008267	['FOUR WISHES' 'DUMB LUCK']
1008279	['NATITA' 'LUNA B']
1008396	['LIGHEA' 'HOLIDAY']
1008413	['MAD SUMMER' 'CYNTHIA']
1008554	['PREDICTION' 'NEW HAMPSHIRE']
1008645	['EL' 'COCO VEINTE']
1008736	['M\\Y BYSTANDER' 'M/Y BYSTANDER']
1008798	['PODIUM' 'C STAR']
1008853	["REGINA D'ITALIA II" "REGINA D'ITALIA" 'REGINA DITALIA' 'MY FAIR LADY']
1008877	['ESCAPE II' 'ESCAPE']

1008889	['LUCY II' 'BAGHEERA']
1008932	['ROSE PIGRE' 'ODYSSEY']
1008994	['ANNA I' 'ANNA']
1009285	['MY PERLE BLEUE' 'M/Y PERLE BLEUE']
1009388	['WHEELS' 'ANASTASIA']
1009508	['ST. DAVID' 'ST DAVID']
1009546	['MARTHA ANN' 'HORIZONS III' 'HORIZONS III']
1009572	['STORMBORN' 'MON PLAISIR']
1009601	['GREYMATTERS' 'GREYM' 'ANDIAMO']
1009742	['LADY NAG NAG' 'LADY BRAVE' 'LADY BRAVE']
1009766	['SYCARA V' 'SYCARA V']
1009792	['MOCHAFY22' 'DYNASTY']
1009807	['ATOMIC' 'ATOM']
1009857	['TACANUYASO M.S.' 'TACANUYASO M.S']
1009869	['MRS L' 'AUSPICIOUS']
1009900	['PRINCESS V' 'M/Y JEREMY']
1009924	['AZTECA' 'AZTEC']
1009948	['NO NAME' 'HURRICANE RUN']
1009962	['AQUILC]' 'AQUILA']
1010117	['VENTUM MARIS' 'INFINITE SHADES']
1010181	['UNITY' 'GRACE']
1010208	['LIND' 'APRIL']
1010260	['TATS' 'SO NICE']
1010284	['PHOENIX2' 'PHOENIX 2' 'P2 SPORTS' 'P2 SEALEGS' 'P2 LIMOSINE']
1010296	['RAASTA' 'NASEEM']
1010337	['HUNTRESS' 'BELLA VITA']
1010416	['NINKASI' 'FABULOUS CHARACTER']
1010959	['ULYSSES' 'BASH']
1011135	['QUANTUM OF SOLACE' 'M/Y VICKY' 'HONOR']
1011238	['SCOUT' 'CALLIOPE']
1011264	['ODYSSEA' 'COMO']
1011408	['PRIDE' 'HEMABEJO' 'HAMABEJO']
1011434	['RIA' 'RAY']
1011496	['DB9' 'DB10']
1011903	['W' 'LARISA']
1012218	['MY ODYSSEY' 'MY CLOUD 9' 'LADY JORGIA']
1012294	['KIBO' 'GRACE']
1012567	["JUST J'S" 'JUST JS']
1012610	['KAOS' 'JUBILEE']
1012737	['LADY LI' 'ANN G']
1012919	['Y717' 'DREAMBOAT']
1012921	['BRAVO Y 718' 'BRAVO EUGENIA']
1012957	['LADY SOFIA' 'AVIVA']
1013066	['OCEANCO - Y717' 'DAR']
1013078	['SATORI' 'DELTA 45']

1013092	['LUNASEA' 'HASNA']
1013145	['MY BOOK ENDS' 'BOOK ENDS']
1230274	['EXUMA' 'DELFINO']
5090414	['TIONEA' 'DIONEA']
8633750	['SEALION' 'SEA LION']
8651099	['RIMA II' 'EGO']
8653310	['LAGNIAPPE' 'BURGAS']
8654601	['SEA EAGLE' 'AXANTHA II']
8660349	["SULLIVAN'S ISLAND" 'SULLIVANS ISLAND' 'MOLIVER']
8662517	['WALLY B-' 'WALINDI']
8666264	['SEADREAM' 'SEA DREAM']
8668286	['SERENDIPITY' 'IMAN']
8745058	['MADCAP' '14 TO SMILE']
8783696	['M.Y. JUSTA DELIA' 'JUSTA DELIA']
8862650	['LOU SPIRIT' 'INSIGNIA']
8933978	['JUST B' 'INTUITION II']
8963997	['MY SOKAR' 'M/Y SOKAR']
8971815	['SS.DELPHINE' 'SS DELPHINE']
8975031	['OHANA' 'DALOLI']
8977479	['ONTARIO' 'LUCY III']
8977534	['IL SOLE' 'IL SCLE']
8979269	['LADY J' 'LADY']
8979714	['KERILEE III' 'KERI LEE III' 'KERI BABY']
8979922	['THEMIS' 'THEMIPARO']
8980311	['M/Y CARPE DIEM II' 'BOUCHON']
8989862	['WEDGETOO' 'WEDGE TOO']
8990677	['MY BACCHUS' 'M/Y BACCHUS']
8994075	['WHEELS I' 'WHEELS']
8996982	['INTUITION LADI' 'EL CARAN']
9097109	['TEMPO' 'MY NANOOK']
9312535	['QUEEN K' 'CLIO']
9350874	['THE LADY K' 'SEVEN SINS']
9378711	['N.M.N' 'ALKHOR']
9385350	['TRENDING' 'TRENDER']
9426180	['PARTY GIRL' 'MILESTONE']
9448621	['BLUE VISN)O' 'BLUE VISION']
9470636	['KATYA' 'AVALON']
9478781	['VICTORY' 'VERTIGO']
9484819	['MY FLOWER I' 'GLADIUS']
9485473	['MY CASINO ROYALE' 'M/Y CASINO ROYALE']
9485485	['THREE FORKS' 'CARTE BLANCHE']
9487017	['BASMALINA2' 'BASMALINA3']
9499785	['ODYSSEY' 'MOON GLIDER' "J'ADE" 'JADE']
9509566	['STAY SALTY' 'HUNTER']
9519080	['ROMANZA' 'CATCHING MOMENTS']

9523067 ['SLO-MO-SHUN' 'FATHOM']
9526758 ['RAYA' 'ONA' 'AL RAYA']
9544633 ['MY SEQUEL P' 'M/Y SEQUEL P' 'M/Y RARE FIND']
9546772 ['ESTEL' 'CELESTIAL HOPE']
9558593 ['MINE GAMES' 'LUMIERE']
9559286 ['KINTA' 'Bella']
9563524 ['EXCELLENCE V' 'ARIENCE']
9570345 ['HELIOS3' 'HELIOS 3']
9570929 ['SEVEN S' 'MANIFIQ']
9586215 ['MY ALAMSHAR' 'M/Y ALAMSHAR']
9586746 ['LAMMOUCHE' 'ALEKSANDAR VII']
9599664 ['LIBERTY' 'LADY SARA']
9600683 ['PLVS VLTRA' 'PLUS ULTRA']
9627679 ['SEPTIMUS' 'MAN OF STEEL']
9631814 ['GALILEO G.' 'GALILEO G']
9633238 ['RELEASE ME' 'ANGEL WINGS']
9633393 ['TESORO' 'ASTERI AYR']
9652868 ['MY ILERIA' 'M/Y ILERIA']
9654921 ['W' 'PIPE DREAM']
9655303 ['M/Y LA PELLEGRINA 1' 'M/Y LA PELLEGRINA']
9658733 ['PHILMX' 'PHILMI']
9668142 ['VANISH' 'HAMPSHIRE']
9679830 ['ILLUSION' 'GALACTICA STAR']
9683154 ["D'NATALIN IV" 'DNATALIN IV']
9692181 ['FRAMURA 3' 'BON VIVANT']
9695274 ['VICA' 'MY VICA']
9709104 ['M/Y AMADEUS' 'BELUGA' 'AMADEUS I' 'AMADEUS']
9710127 ['MY ELEONORA III' 'MY ELEONORA 3' 'LEUDIN I']
9723875 ['BLUSH' 'ARADOS']
9734252 ['ARETI' 'AMATASIA']
9752668 ['ARES' 'ALIVE']
9757761 ['PETRA TARA' 'ENTOURAGE']
9762314 ['PINK SHADOW' 'JOY RIDER']
9777668 ['VOLPINI 2' 'VOLPINI 3']
9794537 ['HALO' 'EJI']
9794575 ['NENINKA' 'AURORA BOREALIS']
9800104 ['TIMELESS' 'FLY ME TO THE MOON']
9803601 ['SYBARIS' 'BADIS']
9812523 ['NARVALO' 'NARVA']
9824825 ['VAN.TOM' 'VAN TOM' 'VAN TOM']
9826196 ['QUEEN D.' 'GITANA']
9840087 ['T/T 2 BYS' 'OCEAN DREAMWALKER 3' 'OCEAN DREAMWALKER 4']
9842906 ['SNOW 5' 'LILIUM']
9848091 ['STELLA MIA' 'STELLA M']
9849198 ['FLYING DAGGER 3' 'FLYING DAGGER']

9861952 ['SEVEN' 'AJ']
9888170 ['STERN' 'CECILIA']

Python code

```
import pandas as pd
import matplotlib.pyplot as plt
import os

def load_and_analyze(file_path, sheet_name):
    df = pd.read_excel(file_path, sheet_name=sheet_name)

    # Count the number of unique IMO numbers and unique ship names
    unique_imo_numbers = df['IMO'].nunique()
    unique_ship_names = df['SHIPNAME'].nunique()

    print("Unique IMO Numbers:", unique_imo_numbers)
    print("Unique Ship Names:", unique_ship_names)

    # Check for vessels with the same name but different IMO numbers
    vessels_diff_imo = df.groupby('SHIPNAME')['IMO'].nunique()
    vessels_with_diff_imo = vessels_diff_imo[vessels_diff_imo > 1]

    print("\nVessels with the same name but different IMO numbers:")
    for shipname in vessels_with_diff_imo.index:
        imos = df[df['SHIPNAME'] == shipname]['IMO'].unique()
        print(f"Ship Name: {shipname}, IMO Numbers: {imos}")

    # Check for IMO numbers with more than one ship name
    imo_diff_names = df.groupby('IMO')['SHIPNAME'].nunique()
    imos_with_diff_names = imo_diff_names[imo_diff_names > 1]

    print("\nIMO numbers with more than one ship name:")
    for imo in imos_with_diff_names.index:
        shipnames = df[df['IMO'] == imo]['SHIPNAME'].unique()
        print(f"IMO Number: {imo}, Ship Names: {shipnames}")

    return df, unique_imo_numbers, vessels_with_diff_imo, imos_with_diff_names

def clean_data(df, vessels_with_diff_imo, imos_with_diff_names):
    # Delete entries with multiple vessel names for the same IMO number
    for imo in imos_with_diff_names.index:
        shipnames = df[df['IMO'] == imo]['SHIPNAME'].unique()
        df = df[~((df['IMO'] == imo) & (df['SHIPNAME'].isin(shipnames)))]
```

```
    # Delete entries with the same name but different IMO numbers
    for shipname in vessels_with_diff_imo.index:
        imos = df[df['SHIPNAME'] == shipname]['IMO'].unique()
        df = df[~((df['SHIPNAME'] == shipname) & (df['IMO'].isin(imos)))]

    return df

def sort_data_by_datetime(df):
    # Combine date and time into a single datetime column for sorting
    df['datetime'] = pd.to_datetime(df['date'].astype(str) + ' ' +
df['time'].astype(str), errors='coerce')

    # Sort the entire dataframe by 'IMO' and 'datetime'
    df = df.sort_values(by=['IMO', 'datetime'])

    return df

def count_marina_visits(df):
    # Initialize a counter dictionary for each marina
    marina_visits = {}

    # Iterate through each unique IMO number
    unique_imos = df['IMO'].unique()
    for imo_number in unique_imos:
        df_yacht = df[df['IMO'] == imo_number].copy()

        # Combine date and time into a single datetime column for sorting
        df_yacht['datetime'] = pd.to_datetime(df_yacht['date'].astype(str) + ' '
+ df_yacht['time'].astype(str), errors='coerce')

        # Sort by datetime
        df_yacht = df_yacht.sort_values(by='datetime')

        # Variables to keep track of previous marina
        previous_marina = None

        # Iterate through the sorted data
        for index, row in df_yacht.iterrows():
            current_marina = row['PORTNAME']
            if current_marina != previous_marina:
                # Count the marina visit
                if current_marina in marina_visits:
                    marina_visits[current_marina] += 1
                else:
```

```
                    marina_visits[current_marina] = 1
            previous_marina = current_marina

    # Sort marina visits in decreasing order
    sorted_marina_visits = dict(sorted(marina_visits.items(), key=lambda item: 
item[1], reverse=True))

    return sorted_marina_visits

def plot_marina_visits(marina_visits, title, filename):
    # Plot the marina visits
    marinas = list(marina_visits.keys())
    visits = list(marina_visits.values())

    plt.figure(figsize=(10, 8))
    plt.barh(marinas, visits, color='skyblue')
    plt.xlabel('Number of Visits')
    plt.ylabel('Marina')
    plt.title(title)
    plt.gca().invert_yaxis()  # Highest values at the top
    plt.tight_layout()
    plt.savefig(filename)
    plt.show()

    # Return the data in table format
    marina_table = pd.DataFrame({'Marina': marinas, 'Number of Visits': visits})
    return marina_table

def save_tables_to_excel(tables, file_path):
    with pd.ExcelWriter(file_path, engine='xlsxwriter') as writer:
        for table_name, table_data in tables.items():
            table_data.to_excel(writer, sheet_name=table_name, index=False)
    print(f"Tables saved to {file_path}")

def print_first_and_last_timestamps(df):
    first_timestamp = df['datetime'].min()
    last_timestamp = df['datetime'].max()
    print(f"First timestamp in the dataset: {first_timestamp}")
    print(f"Last timestamp in the dataset: {last_timestamp}")

def calculate_average_timespan(df):
    # Calculate the time differences between consecutive entries for each vessel
    df['datetime'] = pd.to_datetime(df['date'].astype(str) + ' ' + 
df['time'].astype(str), errors='coerce')
    df = df.sort_values(by=['IMO', 'datetime'])
```

```python
    # Group by IMO and calculate time differences
    df['time_diff'] = df.groupby('IMO')['datetime'].diff()

    # Calculate the average time difference
    average_timespan = df['time_diff'].mean()

    return average_timespan

def main():
    action = input("Do you want to filter and clean data OR run the existing
clean file? (clean/run): ").strip().lower()

    if action == 'clean':
        file_path = 'AIS data results_Jelena clean up - full version.xlsx'
        df, unique_imo_numbers, vessels_with_diff_imo, imos_with_diff_names =
load_and_analyze(file_path, sheet_name='original_zonder IMO')
        cleaned_df = clean_data(df, vessels_with_diff_imo, imos_with_diff_names)
        sorted_df = sort_data_by_datetime(cleaned_df)

        # Save the cleaned and sorted DataFrame back to an Excel file
        cleaned_file_path = 'Cleaned_AIS_data.xlsx'
        sorted_df.to_excel(cleaned_file_path, sheet_name='original_zonder IMO',
index=False)
        print(f"Cleaned and sorted data saved to {cleaned_file_path}")

    elif action == 'run':
        cleaned_file_path = 'Cleaned_AIS_data.xlsx'
        if os.path.exists(cleaned_file_path):
            df = pd.read_excel(cleaned_file_path, sheet_name='original_zonder
IMO')
            unique_imo_numbers = df['IMO'].nunique()
            unique_ship_names = df['SHIPNAME'].nunique()

            print(f"Unique IMO Numbers: {unique_imo_numbers}")
            print(f"Unique Ship Names: {unique_ship_names}")

            # Print the first and last timestamps
            print_first_and_last_timestamps(df)

            # Calculate the average timespan between entries
            average_timespan = calculate_average_timespan(df)
            print(f"Average timespan between two data entries:
{average_timespan}")
```

```
            # Count marina visits for all vessels
            marina_visits = count_marina_visits(df)

            # Generate plots and tables for marinas in sets of 30
            tables = {}
            for i in range(0, 180, 30):
                set_name = f'{i//30 + 1} Set of 30 Marinas'
                marina_set = dict(list(marina_visits.items())[i:i + 30])
                if marina_set:
                    table = plot_marina_visits(marina_set, f'{set_name} Most
Visited Marinas', f'{set_name}_marinas.png')
                    tables[set_name] = table

            # Save the tables to an Excel file
            save_tables_to_excel(tables, 'Marina_Visits_Tables.xlsx')
        else:
            print(f"The cleaned file '{cleaned_file_path}' does not exist.
Please run the clean option first.")
    else:
        print("Invalid option. Please choose 'clean' or 'run'.")

if __name__ == "__main__":
    main()
```