

Analysing AIS Data

Mick Cooney mcooney@describedata.com

Introduction

What is AIS?

AIS ETL

Insurance?

AIS Data

Collision Detection

Mix of dynamic and static data

Automated / Manual

About 30 message types

AIS Message Types

Type 1-2-3 Messages

```
Rows: 136,453
```

Columns: 16

Type 5 Messages

Rows: 127,150

Columns: 17

\$ date	<dtm>	2022-01-01 00:01:49, 2022-01-01 00:14:01, 2022-01-0...
\$ mmsi	<chr>	"311007600", "230091950", "257088370", "305210000", ...
\$ message_type	<chr>	"5", "5", "5", "5", "5", "5", "5", "5", "5", "5", "5...
\$ channel	<chr>	"A", "A", "A", "A", "B", "A", "A", "A", "A", "B", "A...
\$ imo	<chr>	"9073892", "9083885", "9006291", "9057226", "7926409...
\$ callsign	<chr>	"C6XF7@@", "OJRZ", "LAJB8", "V2PU4@@", "OYCM2", "C6X...
\$ name	<chr>	"FRI MARLIN@@@@@@@@", "BALTIC AMELIE", "FALKBRIS",...
\$ shiptype	<chr>	"70", "70", "70", "79", "70", "70", "70", "79", "70"...
\$ dimension	<dbl>	155247046, 161505734, 163619207, 144753030, 14477344...
\$ positional_device	<chr>	"1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1...
\$ eta_month	<dbl>	1, 12, 1, 1, 1, 1, 1, 1, 12, 1, 1, 12, 1, 1, 1, 1...
\$ eta_day	<dbl>	15, 31, 2, 4, 1, 15, 2, 4, 31, 1, 15, 31, 1, 2, 4, 4...
\$ eta_hour	<dbl>	12, 7, 3, 12, 9, 12, 3, 12, 7, 9, 12, 7, 9, 3, 12, 1...
\$...	<chr>	"...", "..."

Data ETL

Different id columns

Vessel: IMO
(7 digits)

Messages: Maritime Mobile Service Identity (MMSI)
(9 digits)

Non-equi joins required

`join_by` in R

`merge_asof` in Python

Further derived values possible

```
Rows: 5,456
Columns: 12
$ imo      <chr> "7926409", "7926409", "7926409", "7926409", "7926409...
$ date     <date> 2022-04-01, 2022-04-02, 2022-04-03, 2022-04-04, 202...
$ year_label <chr> "Year 2022-23", "Year 2022-23", "Year 2022-23", "Yea...
$ ship_stopped <chr> "Stopped", "Not Stopped", "Not Stopped", "Not Stoppe...
$ date_010203 <dtm> 2022-04-01 00:27:59, 2022-04-02 00:11:50, 2022-04-0...
$ mmsi     <chr> "219164000", "219164000", "219164000", "219164000", ...
$ nav_status <chr> "5", "0", "0", "0", "5", "0", "5", "0", "5", "0", "0...
$ turn_rate <dbl> 0, -127, 0, -127, 0, -127, 0, 0, 0, 0, 0, 0, 0, 0, 0...
$ speed_over_ground <dbl> 0, 111, 104, 102, 0, 102, 0, 79, 0, 59, 18, 106, 0, ...
$ position_accuracy <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
$ longitude <dbl> 12.647555, 10.274618, 8.852873, 6.509478, 10.061775,...
$ latitude  <dbl> 55.67678, 57.80967, 57.73769, 58.08997, 57.04983, 56...
```

```
Rows: 5,456
```

Columns: 12

```
$ true_heading      <dbl> 146, 254, 282, 132, 237, 113, 25, 53, 260, 334, 219...
```

```
$ special_maneuver      <chr> "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "...
```

```
$ positional_device <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "...
```

Rows: 5,456

Columns: 14

\$ eta_day	<dbl>	1, 2, 3, 4, 4, 6, 6, 8, 10, 10, 11, 12, 14, 14, 16, 16, ...
\$ eta_hour	<dbl>	8, 1, 11, 23, 23, 18, 18, 19, 19, 19, 1, 20, 19, 16, 8, ...
\$ eta_minute	<dbl>	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
\$ max_draught	<dbl>	49, 71, 49, 72, 72, 49, 49, 72, 48, 48, 48, 64, 48, 48, ...
\$ destination	<chr>	"DKHIR", "DKHIR", "NOREK", "DKAAL", "DKAAL", "DKTHN", "D...
\$ in_geofence	<lgl>	TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TR...
\$ speed_p01	<dbl>	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
\$ speed_p99	<dbl>	129, 129, 129, 129, 129, 129, 129, 129, 129, 129, 129, 129, 129, 1...
\$ draught_p01	<dbl>	42, 42, 42, 42, 42, 42, 42, 42, 42, 42, 42, 42, 42, 42, ...
\$ draught_p99	<dbl>	75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, ...
\$ speed_level	<dbl>	25.8, 25.8, 25.8, 25.8, 25.8, 25.8, 25.8, 25.8, 25.8, 25...
\$ ballast_level	<dbl>	48.6, 48.6, 48.6, 48.6, 48.6, 48.6, 48.6, 48.6, 48.6, 48...
\$ active_status	<chr>	"Waiting", "Transit", "Transit", "Transit", "Waiting", "...
\$...		...

Data Analysis

Ship Utilisation

Load Status

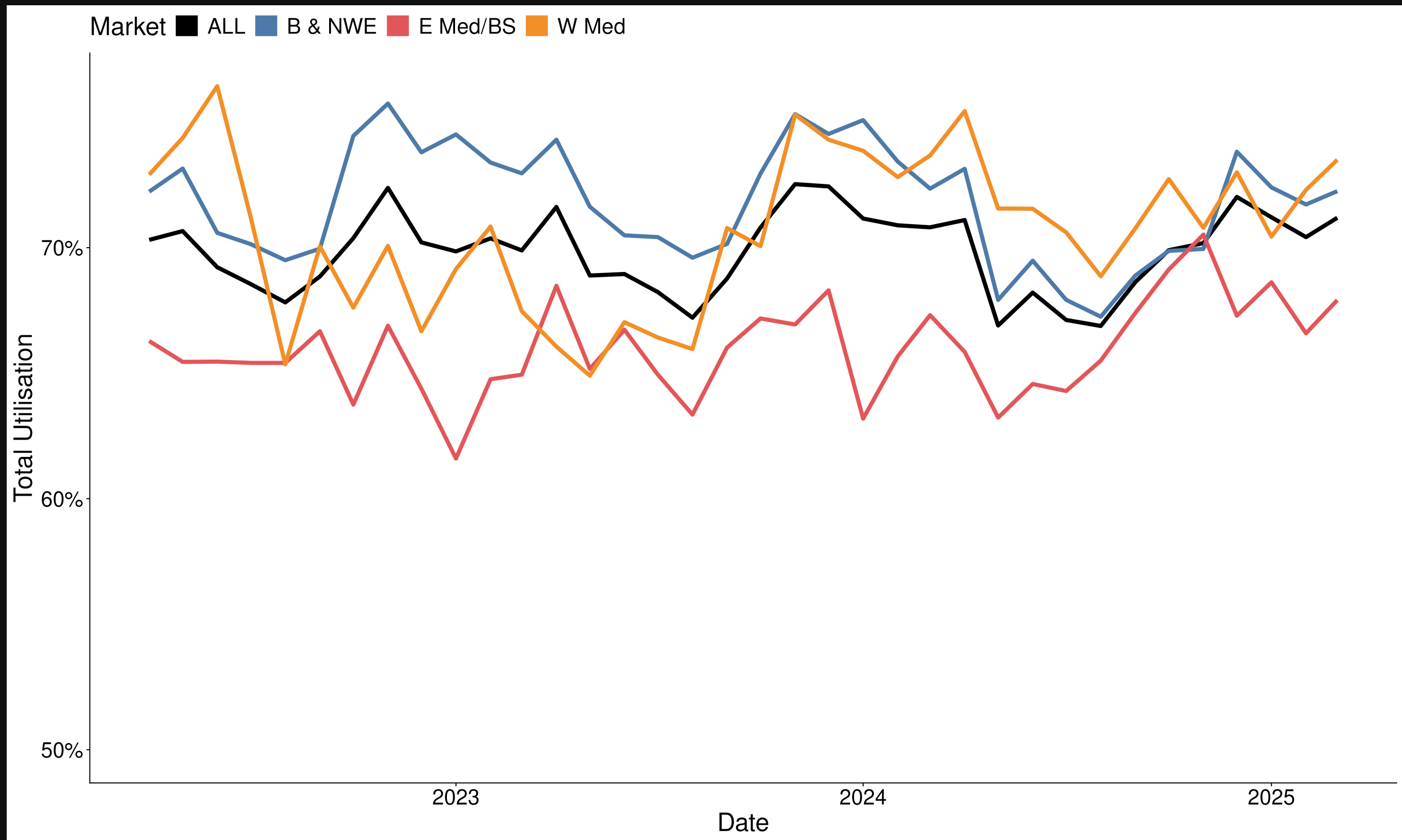
Based on `max_draught` column

Loaded **VS** Ballast

Active Status

Based on ship speed, `speed_over_ground`

`Transit` **VS** `Waiting`



Voyage Analysis

Far from complete

Data quality issues

Interesting preliminary results

```
Rows: 2,696
Columns: 12
$ imo                <chr> "9057226", "9057226", "9057226", "9057226", "9057...
$ section_id         <int> 1, 2, 3, 4, 5, 6, 8, 10, 12, 13, 14, 15, 17, 18, ...
$ voyage_starttime   <dtm> 2022-04-07 14:42:06, 2022-04-11 18:51:53, 2022-0...
$ voyage_endtime     <dtm> 2022-04-10 06:43:01, 2022-04-11 22:49:28, 2022-0...
$ voyage_dist        <dbl> 1140.10, 70.31, 1211.55, 458.97, 0.50, 958.06, 93...
$ voyage_duration_days <dbl> 2.6673032, 0.1649884, 7.2454977, 0.9174884, 1.092...
$ orig_port_fid      <chr> "2232", "3281", "3279", "1435", "3279", "2129", "...
$ orig_port_name     <chr> "TRONDHEIM", "FREDRIKSTAD", "MOSS", "KOBENHAVN", ...
$ orig_port_dist     <dbl> 6.3520539, 19.8100337, 1.3439376, 5.3593236, 1.34...
$ dest_port_fid      <chr> "3281", "3279", "1435", "3279", "2129", "1293", "...
$ dest_port_name     <chr> "FREDRIKSTAD", "MOSS", "KOBENHAVN", "MOSS", "HORT...
$ dest_port_dist     <dbl> 1.7181289, 1.3439376, 2.9690397, 1.3434451, 4.791...
```

Insurance?

Cargo & Specie

Market intelligence

Hull

Thank You!



Slides for Talk