Running for Oil

Finding Emerlad

Florian Gerlinghoff
Student nr.: 2734858
Vrije Universiteit Amsterdam
Amsterdam, Noord-Holland
The Netherlands
f.gerlinghoff@student.vu.nl

Paul Groß
Student nr.: 2776170
Vrije Universiteit Amsterdam
Amsterdam, Noord-Holland
The Netherlands
p.gross@student.vu.nl

Niclas Haderer
Student nr.: 2766114
Vrije Universiteit Amsterdam
Amsterdam, Noord-Holland
The Netherlands
n.haderer@student.vu.nl

ABSTRACT

We analyze trips of crude oil tankers and compare the time it takes them to reach their destination port when the crude oil price is low versus when the price is high. We use a large data set of AIS messages broadcast by ships between 2014 and 2017. The data is filtered, cleaned and transformed in a data pipeline implemented in Apache Spark. The data product of this pipeline is a set of trips of crude oil tankers including their durations. We found that there is a significant difference between trip durations depending on the oil price. It took tankers on average 114.13 hours to reach their destination port when the oil price was high, compared to 142.86 hours when the oil price was low. The results are visualized on a static website.

1 INTRODUCTION

"Buy low, sell high" is a truism not only for stock investors following a market timing strategy, but also for corporations trying to make profit on other markets—like the crude oil market. It does suggest the assumption that, whenever the current oil price is low, the unloading of oil is delayed with the hope that prices will rise again in the near future.

Hypothesis 1. *Crude oil tankers take a longer time for the same route in times when the oil price is low compared to times when the oil price is high.*

We define a *route* as a way across the sea that a ship follows to get from its origin to its destination port. A *trip*, on the other hand, is a journey that a ship makes along a route. In other words, it is a concrete instance of a route.

Thanks to the Automatic Identification System (AIS) which ships use to broadcast their current position, it is possible to track the movement of ships around the globe. This opens up a whole new way to analyze the behavior of ships using a very large data set. But with Big Data also come engineering challenges that require special methods and tools to be solved.

In this paper, we describe the data pipeline we used to transform 2.6 TB of raw AIS messages into a data product and finally a visualization of the trip durations during times of low and high oil prices. Such a visualization helps to collect evidence for or against hypothesis 1.

The rest of the paper is structured as follows. In section 2, we will give some information about AIS and the crude oil price. We give more details on what we consider periods of low or high oil price in section 3. Section 4 describes how we decoded, filtered and transformed the raw AIS data to trips of crude oil tankers from one port to another, and what steps we performed for data cleaning. In section 5, we discuss the findings of the data analysis and how we visualized the results on a static website. Some limitations of our method and implementation are described in section 6. Section 7 summarizes our paper.

2 BACKGROUND

The Automatic Identification System (AIS) was introduced in shipping to aid navigation and help avoid ship collisions by regularly broadcasting information about a ship's status [12]. All ships with a gross tonnage (a measure of internal volume) of 300 tons and more are required to operate an AIS transceiver [11].

A ship sends different messages based on its situation. For example, it is reporting information about its position, speed and heading every few seconds when in motion and every three minutes when at anchor. This information is sent as message types 1, 2 or 3. A message of type 5 is broadcast every six minutes and includes information about the ship itself, including the ship type. The data for this message type has to be entered manually by the crew. A comprehensive overview of other message types can be found at [13].

Every ship is identified by a unique IMO number of the International Maritime Organization, and a unique Maritime Mobile Service Identity (MMSI), issued by a ship's flag state. The MMSI is included in type 1–3 and type 5 messages and can be used for correlating AIS and external data. However, Wu et al. point out that the data might contain duplicated MMSI numbers, as these are entered manually [20].

Thanks to shore-based and satellite-based receivers, live and historic AIS data from around the globe is nowadays readily available on the internet. It has therefore been used in earlier studies. Deng et al. [3] propose a method based on statistical models to remove outliers and other data faults. Furthermore, they enrich the data with information that aid analysis. Kroodsma et al. [9] use AIS to track fish harvesting. They trained neural networks to firstly identify fishing vessels and secondly recognize fishing activity. Yan et al. [21] study the global marine oil trade network and are able to calculate the oil trade volume over sea. Yang et al. [22] survey even more applications of AIS; Svanberg et al. [18] do this with an emphasis on maritime research. Lastly, Tu et al. [19] give an overview of various AIS applications, with a special focus on methods for data mining and analysis.

A multitude of stakeholders participate on the oil market, from oil-producing countries to businesses that consume oil. Their various objectives as well as their reactions to current events influence the price for crude oil. In a short time range, price fluctuations seem to be mainly driven by changes in market expectation, speculation on future markets and political stability in oil-producing countries [6]. In the long-term, other factors come into play, such as the capacity of oil-producing countries [8].

West Texas Intermediate (WTI) and Brent are two benchmarks for crude oil that act as reference prices [17]. Both correlate strongly in the period from 2014 to 2017 (see figure 1). Besides the long-term changes, there are also short-term fluctuations in the price. If we could find that these fluctuations correlate with the trip duration of crude oil tankers, this would be evidence for hypothesis 1.

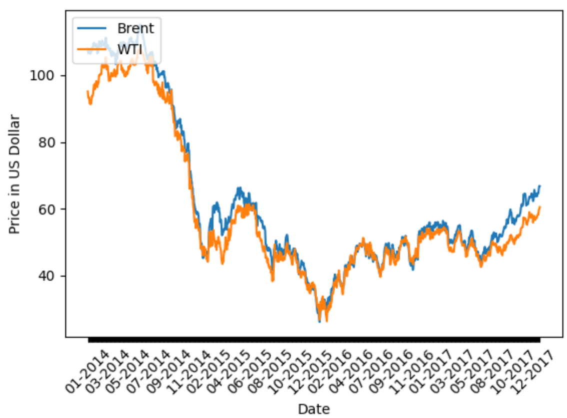


Figure 1: Brent and WTI crude oil price from 2014 to 2017

Figure 2: The oil price

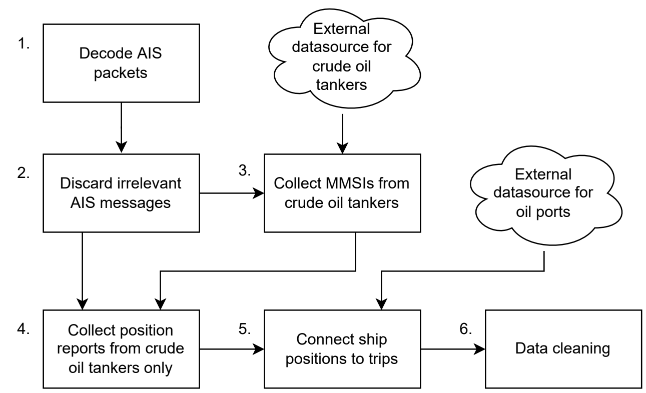


Figure 3: Overview of the stages of our data pipeline

3 OIL PRICE ANALYSIS

To validate hypothesis 1, we have to define periods during which the oil price is considered high or low. In particular, the current change of the oil price is of importance, as this indicates the direction in which the market is moving. When prices are rising, the ships could delay their unloading by loitering to achieve better market timing. When prices are falling, the cargo looses value for each day on sea, so selling as early as possible is financially beneficial. In order to identify periods of rising and falling oil prices, the difference between the ten-day and the 30-day rolling average of the WTI crude oil price is calculated. Then the lower and upper quartile of this difference is defined. Days on which the difference is above the upper quartile are marked as days with rising oil price, while days on which the difference is below the lower quartile are marked as days of falling oil prices. This is visualized in figure 2, with the difference visible as the blue line in the lower graph. We refer to periods consisting of days with rising oil price as high-price periods and to periods consisting of days with falling oil price as low-price periods.

4 DATA PIPELINE

Our data pipeline consists of 6 stages. Firstly, AIS messages have to be decoded. Then they are filtered and the position reports from crude oil tankers are connected to trips. The data is cleaned in a last step.

For the data processing, we use *Apache Spark*[1] on the *Databricks*[2] platform. Spark takes care of the distribution and parallelization of the workload. This allows us to focus on implementing the data processing itself, while the solution is scalable horizontally and vertically. Another benefit of Spark is that we can extend its functionality with user-defined functions (UDFs).

The individual stages of our data pipeline are described in the following sections. Figure 3 visualizes these stages.

4.1 Data Set

The size of our data set is around 2.3 TB. It consists of .txt files that store AIS data packets, with one packet per line. The packets are encoded. Each file contains the packets that were received in one particular minute. The name of a file and the directory it resides in follow the naming scheme ***/yyyy/MM/dd/HH-mm.txt*. Hence, the filename can be used to determine a timestamp for each message.

The data are from October 2014 to August 2017, with some month or days missing in between. Missing data is not a problem, though. It just means that we get less trips in the end, but has no negative consequences for other trips or the analysis itself. The data is mostly from Europe and only from near the shore, where AIS receivers are located.

[2]https://spark.apache.org/
[2]https://www.databricks.com/

4.2 Message Parsing

For the decoding of AIS packets (stage 1), we wrote a Spark UDF. We decided to write it in Scala because this, in general, offers a better performance compared to UDF written in Python [1, p. 111–112]. The UDF passes the raw packets, i.e., lines in the files, to *AISmessages*[3], an open-source Java library, that we use in version 2.2.3. Spark automatically parallelizes this work. However, this approach comes with the drawback that we are not able to decode multi-line packets (see section 6.2).

The result of the parsing stage is a Spark DataFrame which contains the successfully decoded AIS messages from the input files.

4.3 Message Filtering

As described in section 2, AIS transmitters send AIS messages of different types. For the data analysis, we are only interested in position reports (message types 1–3) of crude oil tankers. Ships identify themselves in these messages with an MMSI. To only keep messages from crude oil tankers, we first need to collect the MMSIs of these tankers. For this, we additionally need type 5 messages. Messages of other types than the aforementioned can be discarded (stage 2).

Besides the MMSI, type 5 messages contain an IMO number as identifier, and they provide information on the general ship type as well. A ship type of 80–89 identifies a tanker [13]. But it is not possible to identify *crude oil* tankers from a type 5 message alone.

Therefore, we retrieved a database of approximately 3200 crude oil tankers including their IMO number from VesselFinder[4] through web scraping. We use it to collect all type 5 messages with an IMO that can be found in this database and select the MMSIs from these messages (stage 3). The result is a set of MMSIs from only crude oil tankers that have sent a type 5 message. After processing all files in our original data set, we had a mapping of 354 IMOs from crude oil tankers to their respective MMSI.

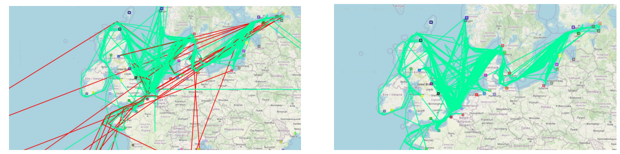
In a next step, we discard all the position reports that do not have an MMSI that can be found in this set (stage 4). This reduces the amount of data by approximately three orders of magnitude. The result of this stage is a DataFrame of 52 million (`MMSI, timestamp, longitude, latitude`) records. This intermediate data product is stored in .parquet files for further processing.

We processed the input files in batches of a three to four month. The processing time for of such a bath was between six and ten hours.

4.4 Trip & Route Creation

Given the position records of crude oil tankers, we connect them to trips in stage 5. The first step in this process is to create a collection of ports visited by each ship in chronological order. To get a list of ports and their geographical location, we used web scraping on three different websites, namely the *Global Energy Observatory*[5], *World Port Source*[6] and *TankTerminals*[7]. We achieved the best results

[3]https://github.com/tbsalling/aismessages
[4]https://www.vesselfinder.com/vessels/
[5]http://globalenergyobservatory.org/
[6]http://worldportsource.com/
[7]https://tankterminals.com/list-of-oil-terminals/



(a) Trips before cleaning **(b) Trips after cleaning**

Figure 4: Result of data cleaning

with the data from TankTerminals, as this website listed only oil ports that are accessible by ship. We merged ports that are not wider than 30 kilometers apart. Each port is the center of a square with an edge length of 15 kilometers. For each position record, if the broadcast location is within such a square, the record was assigned to the corresponding port. This allows us to assemble a chronological list per MMSI of ports that a ship has entered, from which we then create trips between ports.

4.5 Data Cleaning

The trips that are the result of the previous stage are still not yet adequate for a final analysis. In particular, we have to deal with outlier data points and noise. As a final step before the analysis, we perform data cleaning (stage 6).

The result of this stage is displayed in figure 4. In figure 4a, the trips that take a long time are marked in red. These are all successfully removed in figure 4b. Furthermore, most trips do not cross land anymore.

4.5.1 Filtering position updates by coordinates. When we examined the geographical distribution of the positions reported in AIS messages, we found that the majority originated in Europe. However, a few came from the Gulf of Mexico and the cost of Singapore. After visualizing the trips created in stage 5, it became clear that some of them had a discontinuous trace. One possible explanation for this is that ships in two different regions might use the same MMSI. To filter out such faulty position updates within one trip, we narrowed our data down to position reports from Europe only. This was implemented by creating a geo-fence around Europe and removing all AIS position reports which are not within this rectangle. (see figure 5). This measure reduced the number of position reports by 3,908,887 (-7.5%).

4.5.2 Filtering trips by duration. During analysis, trips which took more than 20 days have been found. This is a very long duration for trips within Europe. A reason for such long trips could be that data is missing (i.e., ships do not seem to reach their destination port) or that ships faced unusual events such as urgent repairs. Such trips would have a big impact on the average trip duration in the respective period and introduce a big variance, which makes the statistical evaluation more difficult. Therefore, all trips which took longer than 20 days have been removed.

4.5.3 Moving ports. For a list of positions associated with a ship to become a valid trip, it has to start and end at a port (see section 4.4). If this is not the case, a trip will not be recognized as such and all corresponding data points will be removed. However, as can be seen in figure 6, a lot of ships on their way to St. Petersburg do

Figure 5: Geo-Fencing

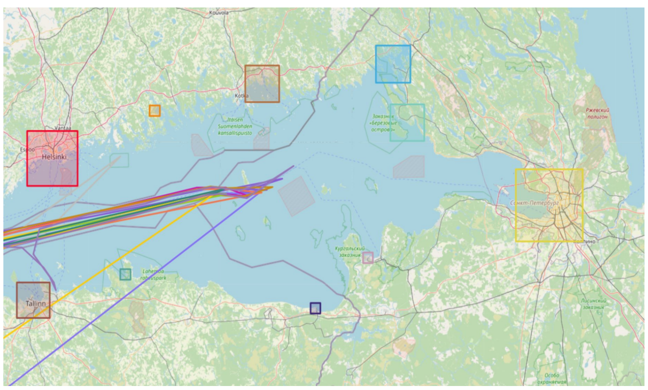


Figure 6: Missing AIS data in Russia

not reach the port. Reasons for this could be, for example, that no data from Russia-based AIS receivers is included in our data set or that ships turn off their AIS transmitter shortly after they cross the Russian border. Since Europe gets most of its oil from Russia (29% in 2020 [4]), we consider this port to be crucial for our data analysis.

Therefore, the positions of the ports have been manually moved to the position where the signal is lost. This is shown in figure 7 and solves the problem, even though it comes with some manual labor. Similar adjustments have been made for the oil port in Immingham, UK.

4.5.4 Velocity gating. As mentioned in section 2, MMSIs can be set manually and are therefore not a very reliable identifier to correlate AIS messages. When two ships have the same MMSI and send position reports at roughly the same time, it seems like a ship jumps across the map. Unrealistic position reports are filtered out using a technique called velocity gating [12]. It is determined if it is physically possible for the ship to reach the reported positions in the time between the messages, with the assumption that crude oil tankers do not usually exceed a speed of 16 knots (29.63 km/h) [21]. If this is not the case, the MMSI is marked as compromised.

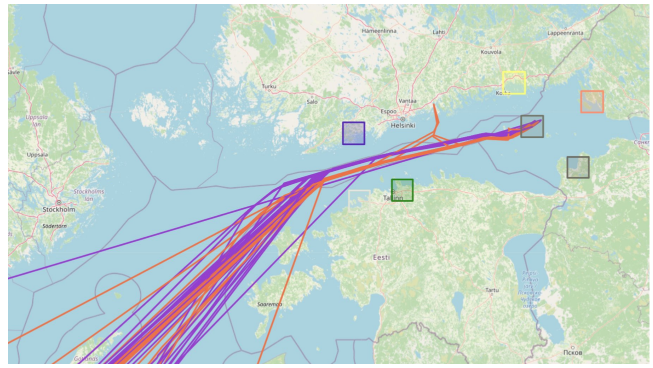


Figure 7: Moving Ports

5 RESULTS

The final data set contains 48,157,636 position reports from 183 different crude oil tankers in the time period from 10/06/2014 to 22/08/2017. From this data, 1708 trips can be created, which can be grouped into 377 routes. Hence, there are 4.53 trips per route on average.

In this section, we will first describe the results of our data analysis and show evidence for hypothesis 1. We will finally describe the visualization of these results.

5.1 Findings

The most traveled route is from Peterburgskiy Neftyanoy Terminal to Rotterdam, which took oil tankers on average six days and was navigated 130 times. More details on the most common routes are shown in table 1.

To answer the key question of this paper—whether ships take longer when oil prices are low—we conducted the following analyses. First, the trips were categorized as happening in a period of either a low, normal or high oil price, according to the time spans depicted in figure 2. If the arrival date was within a high-price or low-price period, the trip was assigned to the high-price or low-price group, respectively. To be able to draw a comparison, we only considered routes that had trips in both the high-price and low-price group. This reduced the total amount of trips included in the analyses by 26.8% to 1250. Furthermore, it is essential to only consider trips of loaded ships which actually intent to sell oil. Therefore, only trips originating in oil exporting countries and ending in non-oil exporting countries are taken into account. As this analysis only covers Europe, the three biggest oil exporting countries are Russia, Norway and the United Kingdom [14]. This further reduced the number of trips to 712 (-43%).

The distribution of the trip durations by price period of the remaining 712 is visualized in figure 8 and listed in table 2. Both average and median seem to support hypothesis 1: During periods of a low oil price, trips tend to take on average 142.86 hours, which is 12.54% longer than during normal periods (126.94 hours). During high-price periods, trips take on average 114.13 hours, which is 10.09% less than the duration during normal periods. The median provides a result which is not as clear the average. With 124.80 hours, the median of high-price periods is only 5.14% higher than

From	To	Trip count	Travel duration [h]
Peterburgskiy Neftyanoy Terminal	Rotterdam	130	148.45
Peterburgskiy Neftyanoy Terminal	Primorsk	92	18.58
Rotterdam	Peterburgskiy Neftyanoy Terminal	92	145.88
Peterburgskiy Neftyanoy Terminal	Ust'-Luga	82	16.26
Kobenhavn	Peterburgskiy Neftyanoy Terminal	81	65.95
Teesport	Immingham	46	27.63
Peterburgskiy Neftyanoy Terminal	Lysekil	43	93.6
Rotterdam	Kobenhavn	38	102.8
Lysekil	Peterburgskiy Neftyanoy Terminal	26	89.25
Rotterdam	Teesport	26	99.31

Table 1: The most common routes sorted by trip count with average travel duration in hours

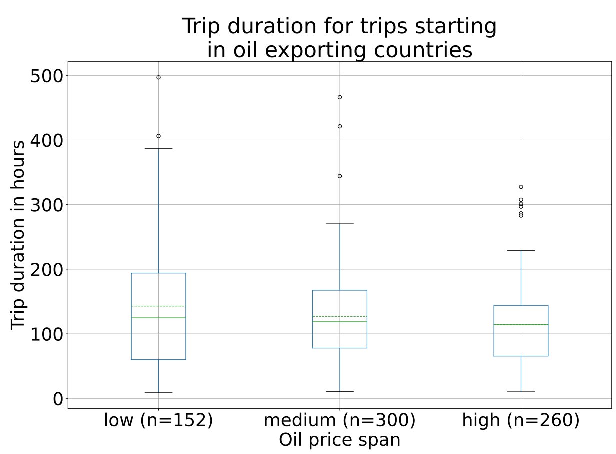


Figure 8: Trip duration in hours by oil price period

Duration	Low	Normal	High
Average [h]	142.86 (+12.54%)	126.94	114.13 (-10.09%)
25% quantile [h]	60.06 (-22.90%)	77.90	65.47 (-15.96%)
Median [h]	124.80 (+5.14%)	118.70	114.15 (-3.83%)
75% quantile [h]	193.96 (+15.85%)	167.42	143.99 (-13.99%)

Table 2: Average, upper and lower quantile, and median trip duration in hours by oil price period with relative difference from values of trips within normal periods

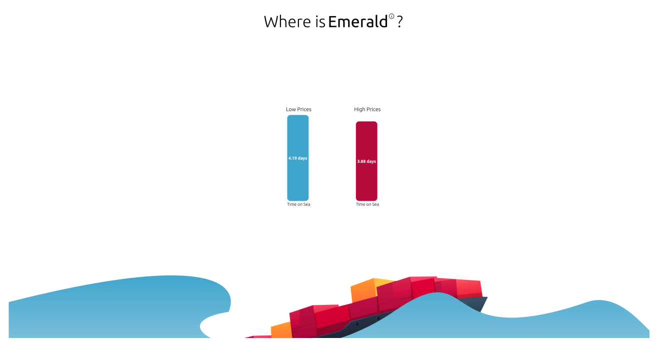


Figure 9: Page 1 – Landing page

the median of normal periods. The median trip duration of low-price periods, too, is with 114.15 only 3.83% smaller than the one normal periods.

As the average is more sensitive to outliers compared to the median and the average shows a clearer result than the median, the reason for the difference between the travel duration of high and low-price periods could be mainly the result of outliers. This offers room for two interpretations: on the one hand, the outliers could be indeed events of loitering and hypothesis 1 would be correct. On the other hand, the outliers could be the result of randomness or noise within the data set.

We performed a paired t-test to evaluate the following two hypotheses:

Null-hypothesis H_0**:** The average duration of trips during high-price and low-price periods is equal, $d_{low} = d_{high}$

Alternative-hypothesis H_1**:** The average duration of trips during low-price periods is greater than the one during high-price periods, $d_{low} > d_{high}$

The result of this test is $t(151) = 4.07, p = 0.93 \cdot 10^{-5}$. As p is smaller than $\alpha = 0.05$, the null-hypothesis H_0 can be rejected. This means that the difference in the duration of trips during periods of high and low oil prices is statistically significant.

5.2 Visualization

Figure 9 displays the landing page of our website. It contains only the most basic information at a glance and allows its visitors to get a quick overview of the results. The diagram in the center displays the average time on sea during periods of a low oil price (blue on the left) and periods of a high oil price (red on the right).

Figure 10 shows all the collected routes. Each of the routes can be selected by clicking on it. After clicking, visitors can see the ship trips that are associated with one route. By default, only a

[7] https://niclashaderer.github.io/oil-ship-tracking/

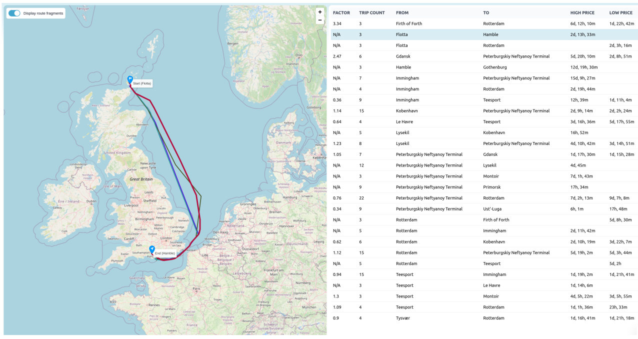


Figure 10: Page 2 – Map with routes

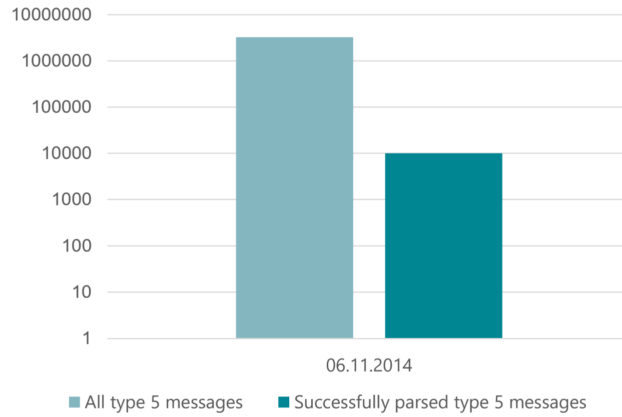


Figure 11: Decoded type 5 messages

pre-calculated “optimal” route will be displayed. By clicking on the button on the top left, additionally the trip fragments extracted from the AIS files are displayed. Because the second page is the part that requires a lot of JavaScript, but is not even visible during the initial page load, the second page gets lazy loaded to improve the website loading time. The same is true for parts of the routes. Only the first ten routes are loaded in by default. Every consecutive batch of ten routes are lazy loaded when the user scrolls the table.

6 LIMITATIONS

Because of the five-week timeline of the project we could not evaluate everything we wanted to and some decisions we made at the start of the project could not be reversed. In this section, we describe some of the limitations[8] to our approach and results.

6.1 Method

With our approach, we can only show a correlation between the oil price and the trip times of crude oil tankers. We did not look into other potential factors.

We also included all trips that took place during high-price or low-price periods, respectively, in the calculation of the average. That is, we did not analyze the difference for every single route on its own. Our results are therefore more susceptible to randomness, such as when tankers incidentally sail a very short route more often when the crude oil price is low, but rarely when it is high.

6.2 Decoding

As mentioned in section 4.2, only AIS messages which do not span more than one line in the input text file will be decoded. We realized this too late and were, due to the constrained time and high utilization of the cluster, not able to get a better approach running in time. In the alternative implementation, we would have to pass entire file streams to the AIS decoder instead of single lines. We struggled with implementing this in Spark while not breaking distribution and parallelization. The one working implementation took a really long time and we decided to interrupt the computation at one point. It was also very hard to debug without separate `stdout` and `stderr` streams per user.

Figure 11 shows an estimation of the number of lost AIS type 5 messages based on data from 06/11/2014. The original data for this day contained 3,266,668 type 5 messages, but our data pipeline was only able to decoded 10,076 of them. This leads us to the conclusion that we only decoded around 0.3% of all type 5 messages successfully. Because AIS type 5 messages are used to determine the identity of the ship as well as the ship type, we most likely lost a substantial amount of data points due to this error.

6.3 Crude Oil Price

We noted in section 2 that there are two big benchmarks for crude oil, which determine their respective crude oil price: Brent and WTI. Even though these prices correlate strongly, they show different short-term fluctuations. For example, political events in the Middle East tend to have a greater impact on the Brent price, while events in the US have a greater influence on the WTI price. For our analysis, we used the WTI price. Perhaps Brent would have been more appropriate because this is the benchmark that is used in Europe—the region where most of our data comes from.

To determine periods of high and low oil prices, we used the 30-day and 10-day rolling average, respectively. This means that, when sudden price changes appear within only a few days and crude oil tankers adjust their behavior because of this, we might not identify these periods as relevant for our analysis.

6.4 Budget

When calculating our budget, we added enough margin for error, and do thus assume that we did stay within our proposed amount. However, we used a lot more computational resources than originally planned. This had mostly to do with not having any experience with a fully utilized Spark cluster. Most of the queries we were running took a lot longer than initially planned because of the high utilization. Another problem was the stability of the cluster. It happened more than once that we had to restart our job because the Spark driver excited unexpectedly or the cluster had to be restarted. As a result, we lost all the data that was processed up to that point. To recude our losses in these events, we started computing smaller batches of data.

[8] What is worst, though, is that we did not find our favorite crude oil tanker Emerald (IMO 9231224) in the data set.

7 CONCLUSION & FURTHER RESEARCH

In this paper, we investigated if the trip duration of crude oil tankers depends on the current crude oil price, such as to achieve better market timing. The hypothesis was that such tankers take longer to reach their destination port when the oil price is low and shorter when the oil price is high. We described a data pipeline, implemented on Apache Spark, that extracted trips of crude oil tankers between oil ports from 2.3 TB AIS messages that were received from 2014 to 2017. A lot of this data we could remove, since we only needed specific messages, namely position reports, from specific ships, namely crude oil tankers, for our data analysis. The remaining positions were connected to a total of 1708 trips.

The statistical test we performed on the trip durations showed that there is a significant difference between oil shipments during periods of a low and high oil prices. This tends to support the hypothesis that oil tankers actually participate in loitering behavior, but it requires future research to draw convincing conclusions.

Researchers (or students) that want to expand on our work can consider various directions. First, our approach was not limitation-free and it would be interesting to see if our results are reproducible by other people when they avoid the mistakes we made (e.g., parsing less than 1% of type 5 messages). Trajectory aggregation [2, 7, 10, 15, 16] could lead to better results during trip and route creation. It would also be interesting to see if the hypothesis still holds when high-price and low-price periods are defined differently, for example with a higher resolution of oil price fluctuations of only a few days. Furthermore, it would be interesting to see if it is possible to "predict" the oil price based on the time ships are on sea. Thanks to satellite-based receivers and more AIS receivers in general, there is much more data available nowadays that also covers the ocean and most parts of the world. This data then also includes the routes between the Middle East and Europe. Future studies can check whether similar results can be found there. Lastly, to get better evidence for loitering behavior, it would be helpful to actually recognize such behavior from the data. Approaches such as the ones described in [5, 23] could be added to the data pipeline.

REFERENCES

[1] Bill Chambers and Matei Zaharia. 2018. *Spark: The Definitive Guide* (1 ed.). O'Reilly, Sebastopol, CA.

[2] Urška Demšar and Kirsi Virrantaus. 2010. Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science* 24, 10 (2010), 1527–1542.

[3] Feng Deng, Sitong Guo, Yong Deng, Hanyue Chu, Qingmeng Zhu, and Fuchun Sun. 2014. Vessel track information mining using AIS data. In *2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*. IEEE, 1–6. https://doi.org/10.1109/MFI.2014.6997641

[4] Eurostat. 2022. From where do we import energy? https://ec.europa.eu/eurostat/cache/infographs/energy/bloc-2c.html Accessed: 28/10/2022.

[5] Dominik Filipiak, Milena Stróżyna, Krzysztof Węcel, and Witold Abramowicz. 2018. Big data for anomaly detection in maritime surveillance: spatial AIS data analysis for tankers. *Zeszyty Naukowe Akademii Marynarki Wojennej* 59 (2018).

[6] Shupei Huang, Haizhong An, Shaobo Wen, and Feng An. 2017. Revisiting driving factors of oil price shocks across time scales. *Energy* 139 (2017), 617–629. https://doi.org/10.1016/j.energy.2017.07.158

[7] Chih-Chieh Hung, Wen-Chih Peng, and Wang-Chien Lee. 2015. Clustering and aggregating clues of trajectories for mining trajectory patterns and routes. *The VLDB Journal* 24, 2 (2015), 169–192.

[8] Peter Y. Jang and Mario G. Beruvides. 2020. Time-Varying Influences of Oil-Producing Countries on Global Oil Price. *Energies* 13, 6 (2020). https://doi.org/10.3390/en13061404

[9] David A. Kroodsma, Juan Mayorga, Timothy Hochberg, Nathan A. Miller, Kristina Boerder, Francesco Ferretti, Alex Wilson, Bjorn Bergman, Timothy D. White, Barbara A. Block, Paul Woods, Brian Sullivan, Christopher Costello, and Boris Worm. 2018. Tracking the global footprint of fisheries. *Science* 359, 6378 (2018), 904–908. https://doi.org/10.1126/science.aao5646

[10] Nirvana Meratnia and Rolf A De By. 2002. Aggregation and comparison of trajectories. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*. 49–54.

[11] International Maritime Organization. 2002. International Convention for the Safety of Life at Sea, 1974 as amended 12/12/2002. http://web.archive.org/web/20041018232014/http://www.navcen.uscg.gov/enav/ais/AIS_Regs_SOLAS_MTSA_FR.pdf Archived: 18/10/2004, Accessed: 30/10/2022.

[12] Giuliana Pallotta, Michele Vespe, and Karna Bryan. 2013. Traffic knowledge discovery from AIS data. In *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 1996–2003.

[13] Eric S. Raymond. 2021. *AIVDM/AIVDO protocol decoding.* https://gpsd.gitlab.io/gpsd/AIVDM.html Accessed: 29/10/2022.

[14] Benjamin Elisha Sawe. 2017. The top oil producing nations in Europe. https://www.worldatlas.com/articles/the-top-oil-producing-nations-in-europe.html Accessed: 28/10/2022.

[15] Roeland Scheepens, Niels Willems, Huub Van de Wetering, Gennady Andrienko, Natalia Andrienko, and Jarke J Van Wijk. 2011. Composite density maps for multivariate trajectories. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2518–2527.

[16] Roeland Scheepens, Niels Willems, Huub van de Wetering, and Jarke J. van Wijk. 2011. Interactive visualization of multivariate trajectory data with density maps. In *2011 IEEE Pacific Visualization Symposium*. 147–154. https://doi.org/10.1109/PACIFICVIS.2011.5742384

[17] Daniel P. Scheitrum, Colin A. Carter, and Cesar Revoredo-Giha. 2018. WTI and Brent futures pricing structure. *Energy Economics* 72 (2018), 462–469. https://doi.org/10.1016/j.eneco.2018.04.039

[18] Martin Svanberg, Vendela Santén, Axel Hörteborn, Henrik Holm, and Christian Finnsgård. 2019. AIS in maritime research. *Marine Policy* 106 (2019), 103520. https://doi.org/10.1016/j.marpol.2019.103520

[19] Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally, and Guang-Bin Huang. 2018. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. *IEEE Transactions on Intelligent Transportation Systems* 19, 5 (2018), 1559–1582. https://doi.org/10.1109/TITS.2017.2724551

[20] Lin Wu, Yongjun Xu, Qi Wang, Fei Wang, and Zhiwei Xu. 2017. Mapping Global Shipping Density from AIS Data. *Journal of Navigation* 70, 1 (2017), 67–81. https://doi.org/10.1017/S0373463316000345

[21] Zhaojin Yan, Yijia Xiao, Liang Cheng, Song Chen, Xiao Zhou, Xiaoguang Ruan, Manchun Li, Rong He, and Bin Ran. 2020. Analysis of global marine oil trade based on automatic identification system (AIS) data. *Journal of Transport Geography* 83 (2020), 102637. https://doi.org/10.1016/j.jtrangeo.2020.102637

[22] Dong Yang, Lingxiao Wu, Shuaian Wang, Haiying Jia, and Kevin X. Li. 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews* 39, 6 (2019), 755–773. https://doi.org/10.1080/01441647.2019.1649315

[23] Zhihao Zhang, Liang Huang, Xin Peng, Yuanqiao Wen, and Lifei Song. 2022. Loitering behavior detection and classification of vessel movements based on trajectory shape and Convolutional Neural Networks. *Ocean Engineering* 258 (08 2022), 111852. https://doi.org/10.1016/j.oceaneng.2022.111852

	Parsing/Filtering	Cleaning	Visualization	Analysis/Evaluation	Report
Florian	60%	0%	0%	0%	33.3%
Paul	20%	50%	50%	100%	33.3%
Niclas	20%	50%	50%	0%	33.4%

Table 3: Work distribution