Time Series Analysis and Prediction of Customers' Invoices

Miikka Kirsilä

School of Science

Bachelor's thesis
Espoo 15.6.2022

Thesis supervisor:

Assoc. Prof. Pauliina Ilmonen

Thesis advisor:

Dr. Tech. Pekka Teppola

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
BACHELOR'S THESIS

Author: Miikka Kirsilä

Title: Time Series Analysis and Prediction of Customers' Invoices

Date: 15.6.2022 Language: English Number of pages: 5+23

Degree programme: Mathematics and systems' sciences

Supervisor: Assoc. Prof. Pauliina Ilmonen

Advisor: Dr. Tech. Pekka Teppola

The increase in competition has led banks to look for new and cost-effective ways to provide a better customer experience. Financial products allow customers to manage their assets, conduct quick payments and pay their bills with unprecedented ease. To find a larger market share and win over customers banks need to find ways to provide better services.

We explore providing the customer with a prediction of their invoice expenditure using time series analysis. This could be especially useful for the customer when they want to understand their monthly spending in a predictive manner. For example, knowing when high-expenditure months are coming might help planning money usage.

More specifically, we use ARIMA and Facebook's Prophet models in addition to a cloud-computing optimized naïve model. These three methods have been chosen not only due to their different approaches but also because of their levels of performance. We primarily use Mean Absolute Percentage Error (MAPE) but also Root Mean Square Error (RMSE) to estimate model accuracy. We also use models with varying history lengths. Namely 6, 12, 24 and 36-months.

After computing models with varying history lengths for 50 real-life customers, we noticed that the naïve approach was the best in terms of MAPE for almost 70% of the customers. It is also by far the fastest to calculate due to the cloud-computing optimized nature.

In the future, the models' performance could be improved in a plethora of ways. We could also use different models such as Artificial Neural Networks.

Keywords: Time series analysis, Finance, Prediction, ARIMA, Prophet

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

KANDIDAATINTYÖN
TIIVISTELMÄ

Tekijä: Miikka Kirsilä

Työn nimi: Time Series Analysis and Prediction of Customers' Invoices

Päivämäärä: 15.6.2022 Kieli: Englanti Sivumäärä: 5+23

Koulutusohjelma: Matematiikka ja systeemitieteet

Vastuuopettaja: Assoc. Prof. Pauliina Ilmonen

Työn ohjaaja: Dr. Tech. Pekka Teppola

Voimistuvan kilpailun vuoksi pankit etsivät uusia ja kustannustehokkaita tapoja tarjota parempaa asiakaskokemusta. Erityyppiset tekniset sovellukset antavat asiakkaille mahdollisuuden hoitaa omistuksiaan, tehdä pikasiirtoja ja maksaa laskujaan ennennäkemättömällä nopeudella. Saadakseen asiakkaita ja suuremman markkinaosuuden pankkien täytyy tarjota parempia palveluita.

Tässä kandidaatintyössä pyrimme ennustamaan henkilöasiakkaiden laskujen maksamiseen kuluvaa rahaa kuukausitasolla. Tämä voisi olla eritoten hyödyllistä asiakkaalle, kun he haluavat ymmärtää tulevaisuudessa tapahtuvaa kuukausittaista kulutustaan. Tästä esimerkkinä voisi olla korkean kulutuksen kuukausien ennustaminen muutaman kuukauden varoajalla.

Lähtökohtaisesti käytämme ARIMA ja Prophet -malleja, joiden lisäksi käytämme kevyempää naivia mallia. Käytämme pääasiallisena mallin tarkkuuden tunnuslukuna keskimääräistä absoluuttista prosenttivirhettä (MAPE), mutta laskemme myös neliöllisen keskiarvovirheen (RMSE). Mainitut kolme mallia lasketaan usealle eri historian pituudelle: 6:lle, 12:lle, 24:lle ja 36:lle kuukaudelle.

Laskettuamme mallit 50:lle oikealle asiakkaalle huomasimme, että naivi malli tuotti parhaan lopputuloksen jopa 70%:lle asiakkaista. Tämä malli on myös kaikista tutkituista malleista nopein, koska se on optimoitu käytössä olevalle pilvipohjaiselle laskennalle.

Tulevaisuudessa mallien toimintanopeutta ja laatua voitaisiin parantaa lukuisin eri keinoin. Ongelmaa voisi lähestyä myös toisilla malleilla, esimerkiksi keinotekoisella neuroverkkolla (ANN).

Avainsanat: Aikasarja-analyysi, Talous, Ennustaminen, ARIMA, Prophet

Preface

I want to thank Assoc. Prof. Pauliina Ilmonen and my instructor Pekka Teppola for the advice and structure they gave to the making of this thesis.

I would also like to thank Sami Niemonen for igniting my interest in mathematics and Juhani Kaila for keeping the flame lit.

Also to those closest to me: Thank you.

Otaniemi, 21.1.2022

Miikka M. A. Kirsilä

Contents

1 Introduction

What if your bank's mobile app could tell you accurately about future spikes in your invoice expenditure? Since private banking has evolved significantly over the decades the chances of accurate data collection and processing have increased. No more do we have to walk into a bank to open a bank account, write checks to transfer money, or fill in invoices for monthly expenses. Online banking has given everybody the possibility to do their banking from home and minimize their time at a physical office.

Even nowadays banks must react to the changing habits of their customers. Mobile phone and broadband subscriptions as global investments to FinTech companies have steadily increased over the beginning of the 21st century providing new challenges to traditional banks.[9] To increase their market share and win over customers banks have to provide better services in the future.

We study a specific case of using data accumulated over long customer relationships to bring value to single customers. To be more specific, we study the usage of historical invoice data with ARIMA and Facebook's Prophet models to give a forecast of future spending. In practice, this could be applied by creating a service where a customer could see past invoices and get a forecast of future invoice expenditure.

Our objective is to find evidence to support or oppose whether these types of calculations can benefit a customer. Due to the rigorous nature of the time series analysis related to the predictions, we also aim to understand whether a simplified or "naive" mathematical model can be applied.

The main question we wish to answer, therefore, is: Is it possible to give accurate enough predictions of invoice spending using time series analysis to benefit private customers?

To be more specific we wish to get answers to the following questions: What type of time series models suit this problem the best? What level of certainty do these models' predictions give and what level can be considered "good enough"? Is it a viable option to use a model which does not require intense parameter estimation and computing?

We use real-life banking data provided by a well-known Finnish bank. Therefore we are conducting an accurate test of time series' analysis capabilities in real-life scenarios.

Situations, where customer payments are forecasted, are not uncommon, but research like this usually focuses on bringing value to a business, not individual customers. For example, newspaper sales forecasting has been done in 2018. [14]

2 Background

Understanding customer behavior is by no means a new practical application of time series analysis since this predictive information can be directly applied to company strategy.

For example in the article [16] Shaohui Ma and Robert Fildes propose a method to implement third-party mobile payments data to forecast their share of daily customer flows. The authors use the newly-developed Gradient Boosting Regression Tree (GBRT) to provide their customers with accurate forecasts.

It is however harder to find articles related to bank customer spending predictions. This could be because banks have the resources to conduct this type of research internally and due to the classified nature of the data do not publish their findings. This thesis also revolves around the customers' needs and bringing value to the customer instead of broadening the bank's insight.

Just like in this thesis, G. Moschini, R. Houssou, J. Bovay, and S. Ropert-Noicoud use in their paper [19] ARIMA models to find fraudulent credit card transactions. Very similarly to this thesis, the authors assume that each customer has a regular spending behavior that is possible to recognize and therefore model. Once this behavior is known, the authors state that "any discrepancies and deviations from it would be likely frauds."

G. Moschini, et al. concluded that despite ARIMA models under-perform, they are still better than other alternatives, such as Box-Plot, LOF, IF, or K-Means. [19]

Facebook's Prophet has been used similarly by Ledion Lico, Indrit Enesi, and Harshita Jaiswal in their paper [18]. The authors analyzed sales in a real-life retail department store in Albania to predict future sales with a low prediction error rate and discover long time trends.

Another interesting application of the Prophet model is with the prediction of COVID-19 -cases. [15] Similar to this thesis, L. Lico, et al. used Mean Absolute Percentage Error (MAPE) - in addition to Mean Absolute Error (MAE) - to assess the performance of their models.

Customer actions have also been modeled using very different methods. A good example of this is a case study where D. Penpece and O. E. Elma used Artificial Neural Networks (ANNs) to model sales in the grocery retailing industry. They managed to get astonishingly accurate results. [7]

3 Research data

The data we use has been provided by a known Finnish bank. The research data is collected from real-life private banking and therefore accurately represents real-life scenarios and situations.

This means that the data is accurate and of extremely high quality. Therefore if a model can adjust to the nature of this collected data, it is very likely that it can also adjust to future situations. On the downside, real-life data is mostly not algorithm-friendly and can prove to be difficult to process.

The data is a table of transactions made by customers since 2018 with each row displaying a single transaction. The columns provided are displayed in more detail below. Also, other information was provided for each transaction, but they are not considered relevant for this thesis.

- Customer name (pseudonymised)
- Account number (pseudonymised)
- Timestamp
- Transaction amount (€)
- Account saldo after transaction
- Customer given name for receiver
- Customer age
- Customer type

The research data contains information through which one could identify customers as well as recipients. To avoid this the customers, accounts, and receivers have been pseudonymized.

The transaction timestamp does not have to be accurate to milliseconds as we are only interested in monthly invoice payments. Because of this, we round each timestamp by the month.

3.1 Data description

The data consists of 4 327 994 rows of individual transactions, where each row represents a single transaction. The data consists of 20 different columns, of which only the relevant ones are used.

The individual transactions have been made by 21 215 different customers. These customers have done the transactions from 29 745 different bank accounts.

In the data, the timestamp is shown with the accuracy of a millisecond, which is too accurate for this thesis. Because of this, we round the transactions to the month of the transaction.

A histogram of the sum of transactions per month can be seen in Figure 1. The Figure takes into account only the transactions in the years 2018, 2019, and 2020, since any years before are not included in the research data and the data from the year 2021 is not yet complete.

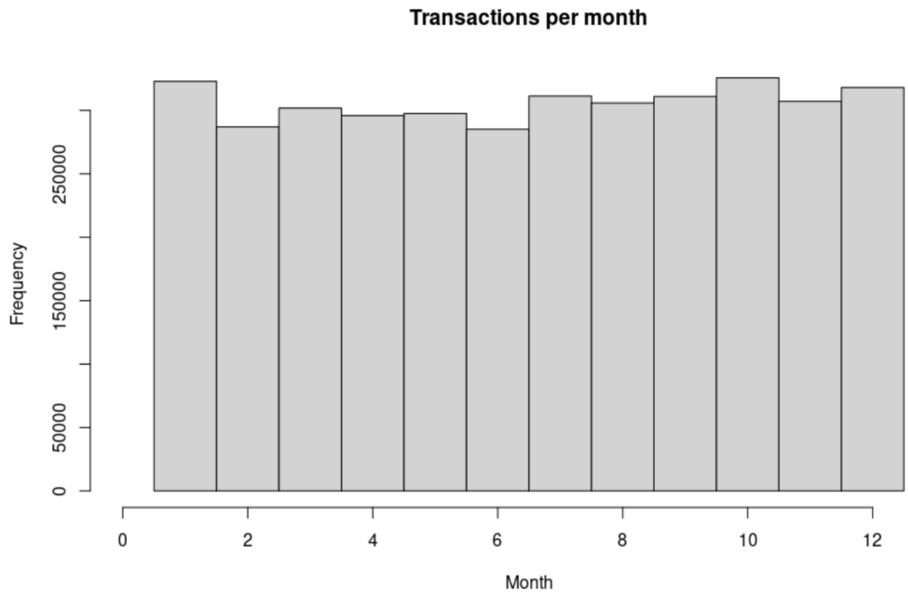

Figure 1: Sum of transactions per month in the years 2018-2020

The amount of transactions gets similar values between different months with no clear exceptions. This could be because most private customers' invoice payments are related to recurring payments that stay rather similar between months. Bills regarding rent, housing, and phone only being a few examples.

In Figure 2 we can see the frequency of invoice payments per day of month. Once again the data is uniform, with only a few exceptions. The first significant exception is on the first day of the month and the second on the 14th day. These come from the monthly and bi-monthly nature of many recurring payments.

The amounts customers transfer in each transaction vary, but the amounts seem to follow a clear distribution. This can be seen in the histogram in Figure 3. The distribution continues beyond the visualization.

The median invoice payment amount is 42.0 € and the mean amount is 144.5 €. Nevertheless, 75% of transactions are under 108.1 €. These values have been calculated from the whole dataset.

The median amount of money remaining on a customer's account after a transaction is 1121.0 € and the mean is 6740.0 €.

In total there are 151 142 differently named recipients. Receiver names are however prone to spelling mistakes and different writing formats since the receiver name is written by the sending party. In this data, receivers are not defined by any other method than the written name. One example of a possible mistake is to write

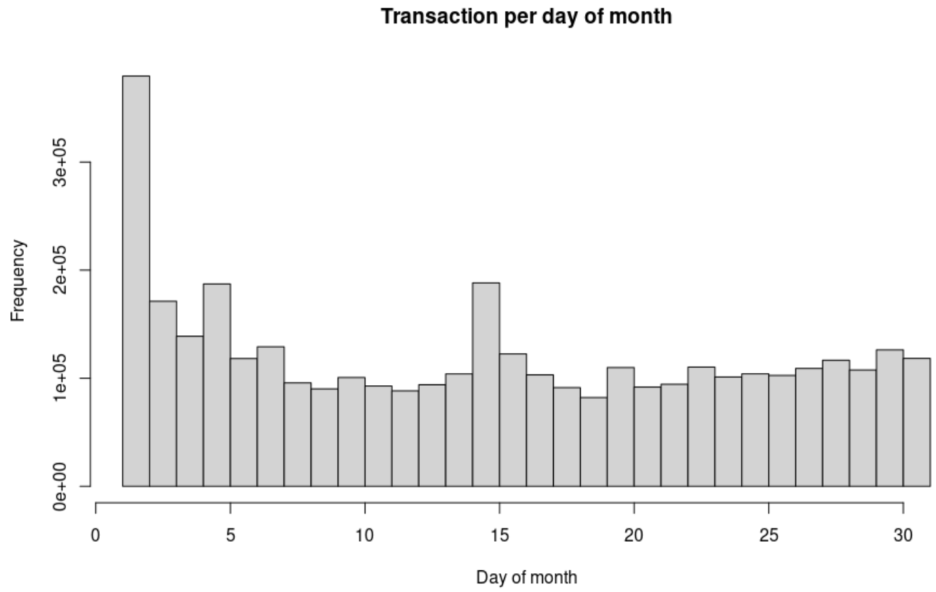

Figure 2: Sum of transactions per day of month in the years 2018-2020

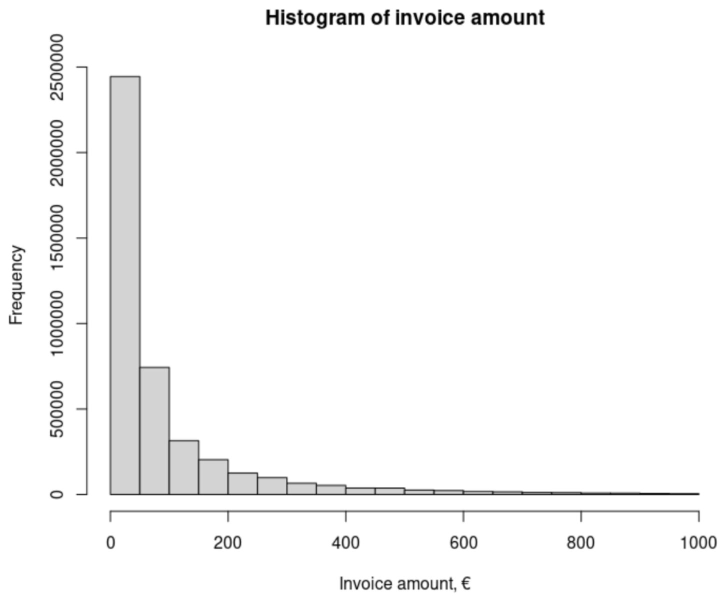

Figure 3: Histogram of transactions per transaction size

either the full name of the company, including the company form, or to leave the company form out. These typos are however irrelevant and we can disregard them.

When we disregard the fact that a single company may be represented by multiple different spellings, the invoice recipients are quite evenly spaced. Nevertheless, there are a few recipients that stand out from the crowd. The 100 recipients with the most received invoices can be seen in Figure 4. For privacy reasons, the recipients' names have been left out.

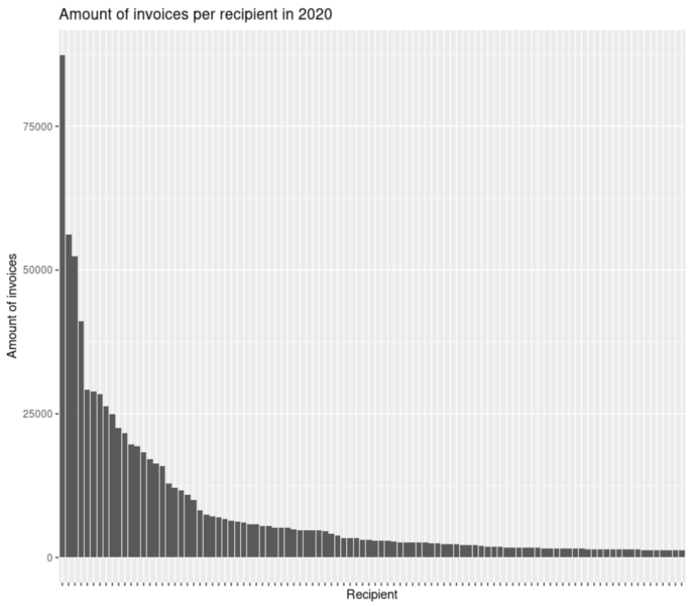

Figure 4: Bar plot of the amount of invoices received by the 100 largest invoice recipients in 2020.

The age distribution of the customers in the year 2020 can be seen in Figure 5. The customer age varies greatly, with every age group being represented. The younger ages however are extremely noticeably present. When sampling the data, the customers were divided into larger segments, and even amounts of customers were picked from each of these segments. Since for example, the youngest segment has the least amount of years in it, it must have more individuals per year to be of equal size compared to the other age segments.

We can also see accounts with suspiciously old owners. These can be for example old estates that have been left unattended. In Finland, the estate inventory must be made, but the estate doesn't have to be split.

Some customer segmentation has also been done beforehand. The customers are split into three different types: "Primary Customer" - a customer who primarily uses this bank, "Partial Customer" - a customer who uses this bank among others, and "Passive Customer" - a customer, who has not been doing a lot of banking for a while. The distribution of these tags from the year 2020 can be seen in Figure 6.

3.2 Selecting data

Due to the very large amount of data acquired from customers, we must limit the data down by a variety of factors. Firstly we are only interested in invoices paid by private customers since we aim to provide a forecasting service to the bank's customers. We can also assume that private customers have a regular spending behavior that can be modeled.

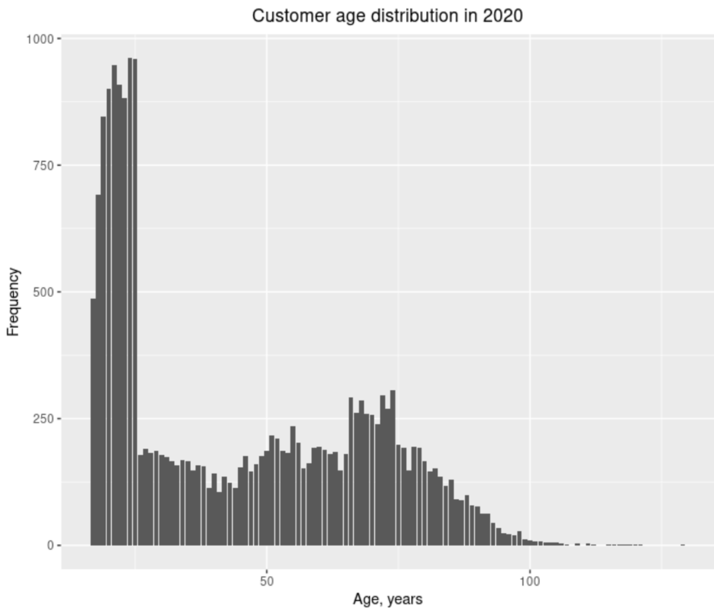

Figure 5: Histogram of the customer age distribution in the year 2020

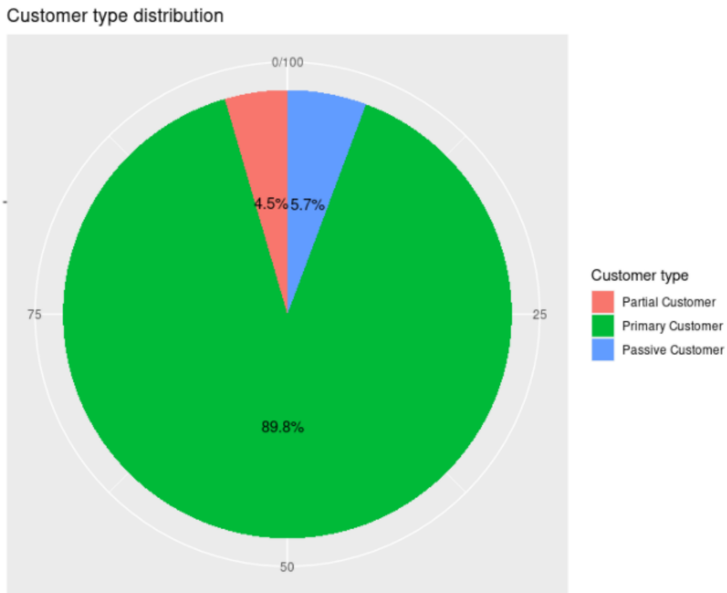

Figure 6: Pie chart of the customer type distribution in the year 2020

The ARIMA model can form a prediction from any amount of history meaning that it can form a prediction for also the 6-month history. When working with monthly data Prophet requires at least 12 months of history to properly model seasonality. This is why we do not use the 6-month history for Prophet.

To compare the prediction to actual historical values we require a single month more. Therefore we must discard all customers with less than 7 for ARIMA or less than 13 months for Prophet of historical data.

These requirements go hand-in-hand with the length of the model history meaning that there are fewer customers with complete 36-month history. After this, we require at least one reference point where to compare the prediction and historical value.

We do not filter data based on the amount transferred, the balance of the account afterward, the receiver, customer age, or customer type, since it is not the aim of this thesis to group customers into segments.

4 Research methods

We use three different approaches to estimate customer invoice spending: the Seasonal ARIMA or SARIMA model, Facebook Prophet, and a more simple approach. SARIMA and Prophet are both well-known models for time series analysis, whereas the naive approach is a very simple way of giving a prediction to give a point to get a reference.

Each of the models is used to produce results from a series of different lengths of history. We have chosen the 6, 12, 24, and 36-month histories to be the ones of special interest. These have been chosen for a few reasons.

Firstly, the research data only spans from the beginning of 2018 to June 2021 so there are in total only 42 months of data. It is not practical to calculate predictions for every length of history, so we chose the few which represent best the multitude of different options.

Let us take a deeper look into the logic behind each of these models.

Brockwell and Davies [5] and James D. Hamilton [1] have good books elaborating on time series analysis. Introductory time series with R [6] is a good introduction to R usage in time series analysis.

4.1 ARIMA models

The ARIMA model or the Autoregressive Integrated Moving Average model can be split into different less complex models. These models are mainly the AR, or autoregressive, and the MA, or moving average piece. These pieces can be further combined with several other operators, such as integration (I).[17]

The AR model is used to forecast a variable of interest using a linear combination of the previous values of the variable in case. The term autoregressive comes from the fact that the model is a regression of the variable itself. [17]

Autoregressive model of order p can be given as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t,$$

where ϕ_1 to ϕ_p are the model parameters and ϵ_t is a white noise error process.

The order of the autoregressive model comes from the number of previous values taken into account. The order p is shown with the AR(p) -notation. Therefore for example a model which uses two previous values of the model would be written as AR(2). [17]

Using a similar method as the AR model, the MA - or moving average - model uses past forecasting errors instead of variable values to provide predictions. In a similar manner as the AR model, the MA model is denoted by MA(q) for the order q. [17]

The MA model can be given as

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q},$$

where $\theta_1, ..., \theta_q$ are the model parameters and ϵ_t is the white noise error process.

To assist the further processing of this model we must add the integrated (I) part to our model. This completes our previous AR and MA models to become an ARIMA model. In case there is a trend in the invoice payments, the integrated model can be used to remove the trend.

The idea behind the integrated part is to generate an ARMA model from the model's first-order difference

$$w_t = y_t - y_{t-1},$$

where y_t is the observed value of the process at time t. These differences can also have versions of a higher orders so that trend from multiple timesteps away can be accounted for. [10]

These differences can be later joined together with the ARMA model. This generates an ARIMA(p,d,q) model where p, d and q are the orders of the AR, I, and MA -models. [10]

Even though this thesis does not make use of these, there are also other ARIMA model derivatives. A few examples of these are Seasonal ARIMA (SARIMA), Gegenbauer autoregressive moving-average processes (GARMA), seasonal autoregressive fractionally integrated moving-average (SARFIMA) model, and flexible seasonal fractionally integrated processes (flexible ARFISMA). [11]

Because these types of ARIMA models are often needed in very different fields of business to generate good enough predictions and due to the lack of manpower to effectively use time series analysis to produce these forecasts, automatic forecasting algorithms have been built. An example of this type of software is for example TRAMO and SEATS, which base their logic on multiplicative seasonal ARIMA model identification. [17]

The ARMA family model parameters can be estimated by using different kinds of iterative methods. See for example [20].

We use the forecast package for R and in particular the functions auto.arima() and forecast() to create time series predictions. auto.arima() provides the optimal ARIMA model orders p, d, and q using unit root tests to the best of its knowledge. The forecast()-function generates future predictions based on the model generated by auto.arima(). [4]

More on the subject of automatic ARIMA model detection can be found from Rob J. Hyndman's and Yeasmin Khandakar's paper Automatic Time Series Forecasting: The forecast Package for R. [4]

4.2 Facebook Prophet

Because of the nature of ARIMA models, small fluctuations in the trend or seasonality at the very end of the acquired data may bring large deviations between the actual data and the prediction of the model. These models also have trouble with understanding the underlying long- and short-term seasonalities since we are not using the SARIMA models.[13]

To combat this we use the Facebook Prophet package suggested by Sean J. Taylor and Benjamin Letham in their article [13]. The Prophet package suggests an easy-to-understand model, with "intuitive parameters that can be adjusted without knowing the details of the underlying model."

Taylor and Letham propose a decomposable time series model

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t),$$

which contains three components: trend $g(t)$, seasonality $s(t)$ and holidays $h(t)$.[13]

Let us begin by further understanding the innermost workings of the trend $g(t)$. To better suit the applications of Facebook, the Prophet model is a combination of two trend models: a saturating growth model and a piecewise linear model. Both of these models are complex so it isn't in the scope of this thesis to deeply understand them. [13]

Prophet relies heavily on Fourier series to account for time series' seasonality $s(t)$. These standard smooth seasonalities can be approximated reasonably well by the Fourier series

$$s(t) = \sum_{n=1}^{N}(a_n cos(\frac{2\pi nt}{P}) + b_n sin(\frac{2\pi nt}{P}))$$

. Here P is the length of a certain seasonality, for example, $P = 7$ for weekly data or $P = 365.25$ for yearly data. The parameters $\beta = [a_1, b_1, ..., a_N, b_N]^T$ are still required to create the seasonal component $s(t) = X(t)\beta$. [13]

This can be done by creating a matrix of seasonality vectors for t in our historical and future data. An example of this is

$$X(t) = [cos(\frac{2\pi(1)t}{265.25}), ..., sin(\frac{2\pi(10)t}{365.25})],$$

with $N = 10$. [13]

Taylor and Letham state that "holidays and events provide large, somewhat predictable shocks to many business time series and often do not follow a periodic pattern, so their effects are not well modeled by a smooth cycle." For example, Thanksgiving in the United States only occurs on the fourth Thursday of every November. [13]

To better account for this, the Prophet model generates a matrix of regressors much like in the processing of seasonality. This is demonstrated by stating that

$$Z(t) = [1(t \in D_1), ..., 1(t \in D_L)]$$

and

$$h(t) = Z(t)\kappa,$$

where $h(t)$ is the holiday component and the change in forecast corresponding to each holiday $\kappa \sim Normal(0, v^2)$.

The Prophet package for R does provide some country-specific holidays, but since there aren't any pre-made dates for Finland in specific, we found this part to be out of the scope of this thesis. Further investigation could be done, but since national holidays are not the only days that could affect private customer expenditure, this task could prove arduous.

4.3 The naive approach

To better understand the goodness of the predictions SARIMA and Prophet provide us we also compare them to a simple, "naive" approach. The naive approach is designed to be a very efficient model which is easily implemented into the cloud computation currently in use.

In practice, the naive model forms its one-month prediction by calculating the simple average of the previous n months. In theory, this means that the naive model is only an AR model which gives the same weight to each of the historical elements.

To justify the significantly longer computation times Prophet and SARIMA should at least yield better results than this naive approach.

4.4 Estimating the goodness of the prediction

For this thesis, we are using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) to define the goodness of the prediction.

MAPE was selected due to its relative nature, where it gives results that can be compared even between extremely different scale values. This is very necessary due to the nature of the data. In some cases, the sum of a single month's transactions might be a six-figure number whereas in some cases we might be talking about mere tens of euros.

MAPE is calculated with

$$MAPE = mean(|p_t|),$$

which can also be seen as taking the simple average of absolute percentage errors. Here p_t is the percentage error, which comes from $p_t = \frac{100e_t}{Y_t}$.[3]

RMSE on the other hand is denoted by

$$RMSE = \sqrt{mean(e_t^2)},$$

which is the square root of the mean of the scaled prediction error.[3]

RMSE is seen as relative to this thesis because of the way it gives weight to large differences between the predicted and actual values. Customers with very large and very small monthly transaction sums are not however comparable by RMSE.

There are also other methods to define prediction goodness. One good example is Symmetric Mean Absolute Percentage Error (SMAPE). SMAPE is calculated by comparing the absolute error between the prediction and actual value to the mean of the prediction and actual value. SMAPE is useful in scenarios where it can be said that the actual value is not completely accurate, for example when data is acquired with measurement instruments. In our case, the acquired values are based on real customer data and are completely accurate so SMAPE cannot be used.

Another example of the practical application of MAPE and RMSE can be found from in [8].

5 Results

On the first round of mass testing, we generated the naive, ARIMA, and Prophet models for 50 randomly selected customers to limit processing time. Each of the models was generated using 6, 12, 24, and 36 months of previous data. Some summary statistics can be seen in Table 1.

The results seen in Table 1 should be interpreted as follows: The closer the MAPE is to zero, the better the result. The values are also prosentual shown in decimal numbers. This means that for example the number 0,82 represents the number 82 %.

As previously stated, we also calculate the RMSE which is an absolute variant of MAPE. They both depict the same thing, but RMSE compares the absolute invoice sums whereas MAPE is relative. The RMSE values are visible in Table 2, but because of their absolute nature are not relative in this thesis' scope and as such are not shown for the rest of the thesis. The smaller the RMSE, the better.

Table 1: Median and Mean of the MAPE-values of the ARIMA, Prophet and naive models with a test group of 50 persons on the second round of testing. Numbers rounded to five decimals.

	6 months	12 months	24 months	36 months
ARIMA median	0,82105	0,82054	0,80379	0,75718
Prophet median	NA	1,1526	0,7273	0,6135
naive median	0,9888	0,8939	0,4339	0,3525
ARIMA mean	1,48626	1,60351	1,27212	0,95904
Prophet mean	NA	2,19	2,029	1,5028
naive mean	2,0704	1,5212	0,9621	0,4995

Table 2: Summary of the RMSE-values of the ARIMA, Prophet and naive models with a test group of 50 persons on the second round of testing. Numbers rounded to three decimals.

	6 months	12 months	24 months	36 months
ARIMA median	655,107	659,29	688,536	453,787
Prophet median	NA	399,7	496,9	448,5
naive median	492,2	449,4	328,4	249,5
ARIMA mean	1200,349	1187,649	1161,505	930,699
Prophet mean	NA	984,5	1085,3	711,3
naive mean	1123,5	1019,7	691,3	403,9

The Tables 1 and 2 contain NA -values on the 6-month Prophet predictions. This is due to the Prophet model creating its predictions - amongst other things - based on weekly, monthly, and yearly seasonality. As the research data is monthly, the closest possible seasonality exists on a yearly level. The 6-month history does not have enough data for a model of this scale.

An interesting note is that the naive model gets drastically different results when compared to the ARIMA model even though it is only a certain type of AR model. We suspect that this is due to imperfections in the auto.arima() -tool which we use to define the ARIMA model parameters.

5.1 Best prediction

Even though by median and mean the naive model with a 36-month-history seems to be the best option for the whole population, this might not be the case for every individual customer.

To achieve the results displayed in Table 1 we have calculated each of the three models with each of the four history lengths for each customer. Let us now choose the best model (naive, ARIMA, or Prophet) for each customer history length. The best model is here the model with the smallest MAPE on a certain customer and history length. The summary of these values can be seen in Table 3.

Table 3: Summary of the MAPEs when the best model is chosen per individual customer.

Training data	min	median	mean	variance	max
6 months	0,04977	0,82105	1,48626	3,694194	11,40406
12 months	0	0,6381	0,9959	1,114902	4,9169
24 months	0	0,6336	1,0285	1,375415	6,2517
36 months	0	0,5243	0,7761	0,6219011	3,3688

The histograms containing the amounts of different MAPE-values for each of the different models can be found in the Figures 7, 8 and 9. These figures may contain multiple entries per customer, one for each length of history.

Now, these results do not contain the whole truth either, so for the final step, we calculate the optimal history length for each of the best models. After doing this we end up with the data summarized by Table 4.

Table 4: Summary of the MAPEs when the best model per individual customer is combined with optimal history length per customer.

	min	median	mean	variance	max
Combined best months	0	0,3486	0,4718	0,1065467	1,6043

A histogram containing the amounts of the different MAPEs, when the best

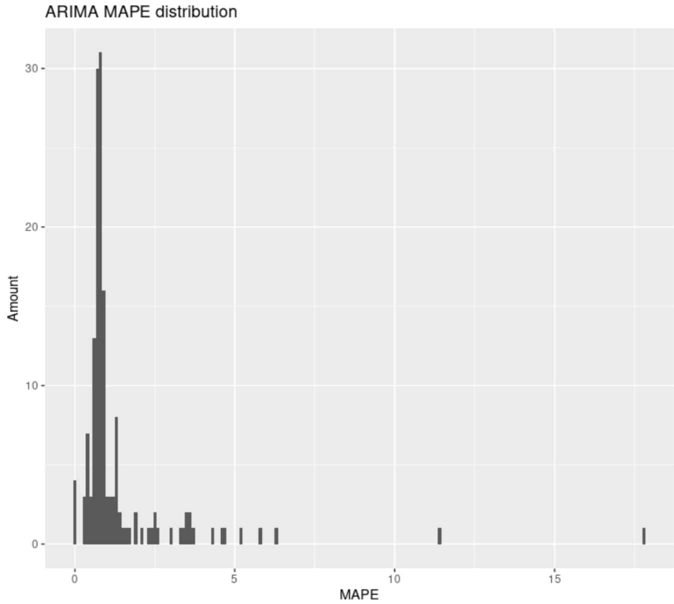

Figure 7: Histogram of the MAPE-values of the ARIMA model predictions.

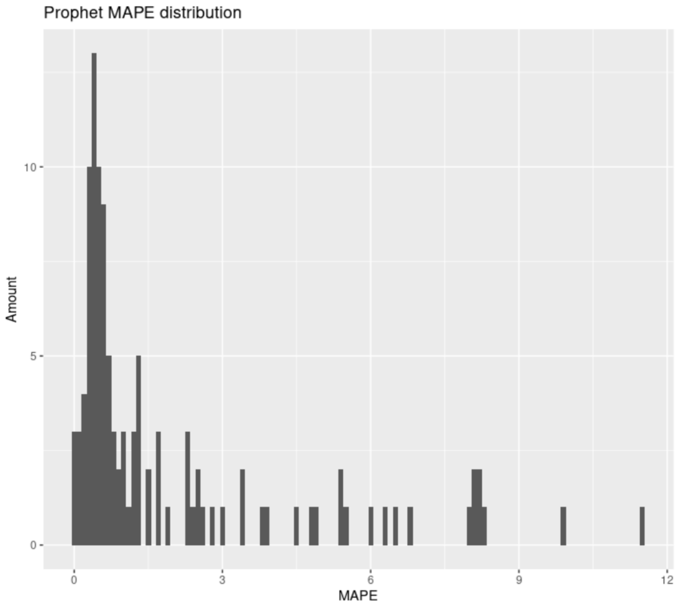

Figure 8: Histogram of the MAPE-values of the Prophet model predictions.

model has been chosen both by history length and by model can be seen in Figure 10. This histogram represents also the values shown in Table 4.

The models delivered the following amounts of best predictions shown in Figure 11. The Figure shows clearly that the naive approach provided the largest amount of predictions with the lowest MAPE.

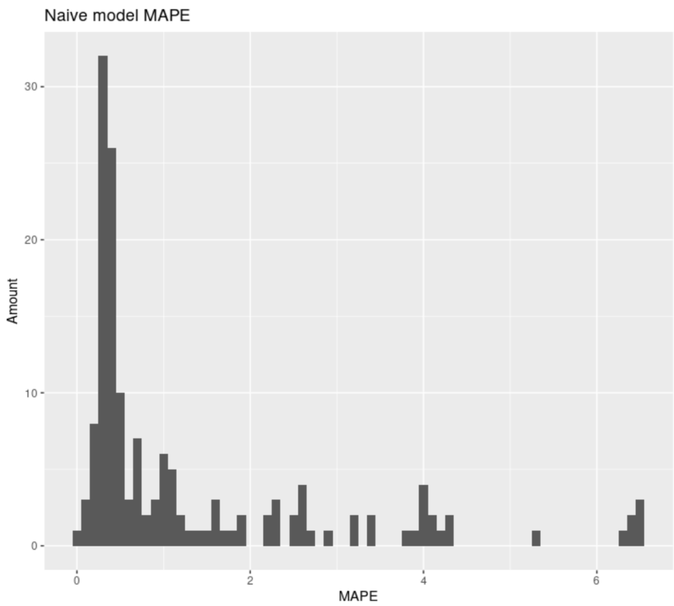

Figure 9: Histogram of the MAPE-values of the naive model predictions.

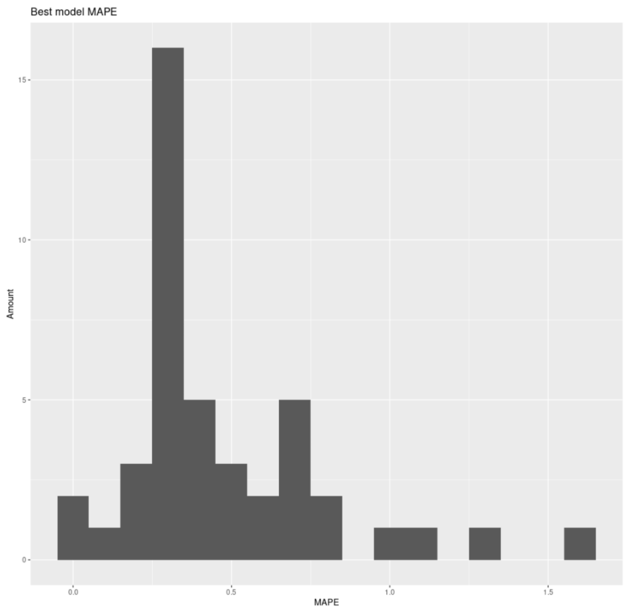

Figure 10: Histogram of the MAPE-values when the best model is chosen from a set of different training data lengths and the three models.

5.2 The worst prediction

The worst prediction, which is shown in Tables 5 and 6, has been obtained using the same method as obtaining the best prediction.

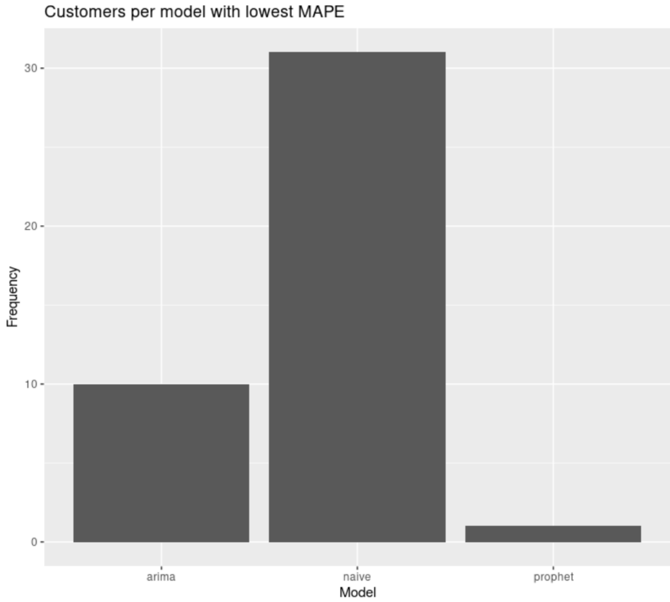

Figure 11: The amount of best predictions by model.

Table 5: Summary of the MAPEs when the worst model is chosen per individual customer.

Training data	min	median	mean	variance	max
6 months	0,04977	0,82105	1,48626	3,694194	11,40406
12 months	0,04977	1,15265	2,79763	11,91872	17,76007
24 months	0,04977	0,87603	2,27264	6,950674	11,45
36 months	0,04977	0,81704	1,68574	4,218728	9,86518

5.3 Examples

To better visualize the models we have selected a few interesting customers and drawn their monthly spending patterns. The visualizations can be found in appendix 1. Table 7 shows the customer id and why this customer is interesting. All of these customers are a part of the sample group used by the models.

Table 6: Summary of the MAPEs when the worst model per individual customer is combined with optimal history length per customer.

	min	median	mean	variance	max
Combined best months	0,04977	0,73311	1,17391	3,093816	11,40406

Table 7: What different customers are examples of.

The most successful models	18399, 27627, 8523, 28510
The worst models	581, 11939, 13877, 13450
Good Prophet models	27627, 18399
Bad Prophet models	11939, 105990
Good ARIMA models	27627, 64907
Bad ARIMA models	581, 11939
Least difference between models	10300, 10334
Most difference between models	581, 28682

6 Future prospects

Since all of the used models underperformed significantly we can directly deduce that major changes to the assumptions and system of prediction must be made.

Currently, this thesis does not make use of Prophet's capability to take into account different holidays. Since we are talking about a Finnish bank, it would be relatively easy to list dates, which could affect customer spending.

Customers could be split into different groups using for example K-Means clustering proposed by A. Likas, et al. [2] After this we could try to form customer-group-based predictions.

We used auto.arima() -library, which does not provide the optimal solutions to finding the best possible ARIMA model. This is clear especially when comparing the ARIMA model results to the significantly better naive model results. Other methods of ARIMA parameter estimation should be explored.

Also, methods with Artificial Neural Networks (ANN) could be useful and provide good results. ANN models have been previously compared to ARIMA models with varying results. For example when forecasting electricity prices the ARIMA(4,1,2) model gave better results in terms of RMSEs of price forecasts than the ANN (20 neurons, 4 delays) model. [12] In comparison, sales modeling in the grocery retailing industry has previously been extremely successful. [7]

The payments of a single customer could also be grouped by receiver name and generate the forecast on a receiver basis. This model would be based on the assumption that customer-receiver interactions follow a certain pattern that could be modeled. This is a different assumption than the one we are working with right now. Currently, we assume that the complete customer invoice behavior is modellable.

7 Summary

The modeling of customer spending activity is by no means a new practical application of time series analysis. As presented in this thesis, multiple methods ranging from ARIMA models and Facebook's Prophet to artificial neural networks have been presented when crafting case studies of anything from grocery stores to epidemiological analysis. Meanwhile, banks are struggling with increased competition and wish to find new and cost-effective methods to provide improved customer experiences.

We studied the use of time series analysis to provide accurate forecasts of monthly customer invoice expenditure data. This data could later be used to visualize expenditure and create alerts of incoming months with large invoice costs.

The specific time series' models we decided to use were ARIMA and Facebook's Prophet due to their feasibility in this thesis' scope and library availability. To combat the inefficiency of these models we took a simple "naive" model to act as a reference point when comparing models. The naive model is a simple average of select length of history and therefore adapts well to cloud computing.

As defined in the introduction, the main question of this thesis was: "Is it possible to give accurate enough predictions of invoice spending using time series analysis to benefit private customers?"

To be more specific, we divided this question into the following subproblems: What type of time series models suit this problem the best? What level of certainty do these models' predictions give and what level can be considered "good enough"? Is it a viable option to use a model which does not require intense parameter estimation and computing?

The time series model that best suits this problem seemed to be the naive model. Even though it is only a simple AR model, it worked surprisingly well. It outperformed both the ARIMA and Prophet models and for over 70% of our test group's customers some history length of the naive model provided the best results.

We define the best model by the MAPE-values it gives out when used on different real-life customers. We also give emphasis on the naive model, because of it's high performance and optimized way of working.

The level of certainty the best models gave was a median MAPE of 35%. In practice, this means that the models' predictions deviated by 35% from the actual result at median. Whether or not this is satisfactory is left for the reader, but we must remember that there have also been customers who fall below this median line.

Even though we did not do performance testing, we know that the simple average calculation of the naive model is by far a faster process than the forming of an optimal ARIMA or Prophet model. Since the naive model outperformed both of its competitors, we can say that it is a viable option.

From all the data gathered during this thesis, we can state that the naive model can be used in a well-performing manner to achieve accurate predictions for a portion of this bank's customers. This gives us reason to further explore the use of these models in the future.

After sufficient explorations these predictions could be productised to bring value to banks' customers.

References

[1] J.D. Hamilton and Princeton University Press. *Time Series Analysis*. Time Series Analysis nid. 10. Princeton University Press, 1994. ISBN: 9780691042893. URL: https://books.google.fi/books?id=B8%5C_1UBmqVUoC.

[2] Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek. "The global k-means clustering algorithm". In: *Pattern Recognition* 36.2 (2003). Biometrics, pp. 451–461. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(02)00060-2. URL: https://www.sciencedirect.com/science/article/pii/S0031320302000602.

[3] Rob J. Hyndman and Anne B. Koehler. "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4 (2006), pp. 679–688. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0169207006000239.

[4] Yeasmin Khandakar Rob J. Hyndman. "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software* 27.3 (2008). DOI: http://dx.doi.org/10.18637/jss.v027.i03.

[5] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer New York, 2009. ISBN: 9781441903204. URL: https://books.google.fi/books?id=TVIpBgAAQBAJ.

[6] Paul SP Cowpertwait and Andrew V Metcalfe. *Introductory time series with R*. Springer Science & Business Media, 2009.

[7] Dilek Penpece and Orhan Emre Elma. "Predicting sales revenue by using artificial neural network in grocery retailing industry: a case study in Turkey". In: *International Journal of Trade, Economics and Finance* 5.5 (2014), p. 435.

[8] The PLOS ONE Staff. "Correction: Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data". In: *PLOS ONE* 9.2 (Feb. 2014), pp. 1–1. DOI: 10.1371/journal.pone.0091629. URL: https://doi.org/10.1371/journal.pone.0091629.

[9] Carmen Cuesta et al. "The digital transformation of the banking industry". In: *BBVA research* (2015), pp. 1–10.

[10] Cheryl L. Jennings Douglas C. Montgomery and Murat Kulahci. *Introduction to Time Series Analysis and Forecasting*. John Wiley Sons, Incorporated, 2015. ISBN: 9781118745229.

[11] Wilfredo Palma. *Time Series Analysis*. John Wiley Sons, Incorporated, 2016. ISBN: 9781118634349.

[12] Gao Gao, Kwoklun Lo, and Fulin Fan. "Comparison of ARIMA and ANN models used in electricity price forecasting for power market". In: *Energy and Power Engineering* 9.4B (2017), pp. 120–126.

[13] Letham B. Taylor SJ. "Forecasting at scale". In: *PeerJ Preprints 5:e3190v2* (2017). DOI: https://doi.org/10.7287/peerj.preprints.3190v2.

[14] Carina Intan Permatasari, Wahyudi Sutopo, and Muh. Hisjam. "Sales forecasting newspaper with ARIMA: A case study". In: *AIP Conference Proceedings* 1931.1 (2018), p. 030017. DOI: 10.1063/1.5024076. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.5024076. URL: https://aip.scitation.org/doi/abs/10.1063/1.5024076.

[15] M Indhuja and P Sindhuja. "Prediction of covid-19 cases in India using prophet". In: *Int. J. Stat. Appl. Math* 5.4 (2020).

[16] Shaohui Ma and Robert Fildes. "Forecasting third-party mobile payments with implications for customer flow prediction". In: *International Journal of Forecasting* 36.3 (2020), pp. 739–760. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2019.08.012. URL: https://www.sciencedirect.com/science/article/pii/S0169207019302365.

[17] George Hyndman Rob J; Athanasopoulos. *Forecasting: principles and Practice*. oTexts, 2021. ISBN: 9780198520115. (accessed: 03.08.2021).

[18] Ledion Liço, Indrit Enesi, and Harshita Jaiswal. "Predicting Customer Behavior Using Prophet Algorithm In A Real Time Series Dataset". In: (2021).

[19] Giulia Moschini et al. "Anomaly and Fraud Detection in Credit Card Transactions Using the ARIMA Model". In: *Engineering Proceedings* 5.1 (2021). ISSN: 2673-4591. DOI: 10.3390/engproc2021005056. URL: https://www.mdpi.com/2673-4591/5/1/56.

[20] Gidon Eshel. *The Yule Walker Equations for the AR Coefficients*. URL: http://www-stat.wharton.upenn.edu/~steele/Courses/956/Resource/YWSourceFiles/YW-Eshel.pdf.

Appendix 1: Sample customers

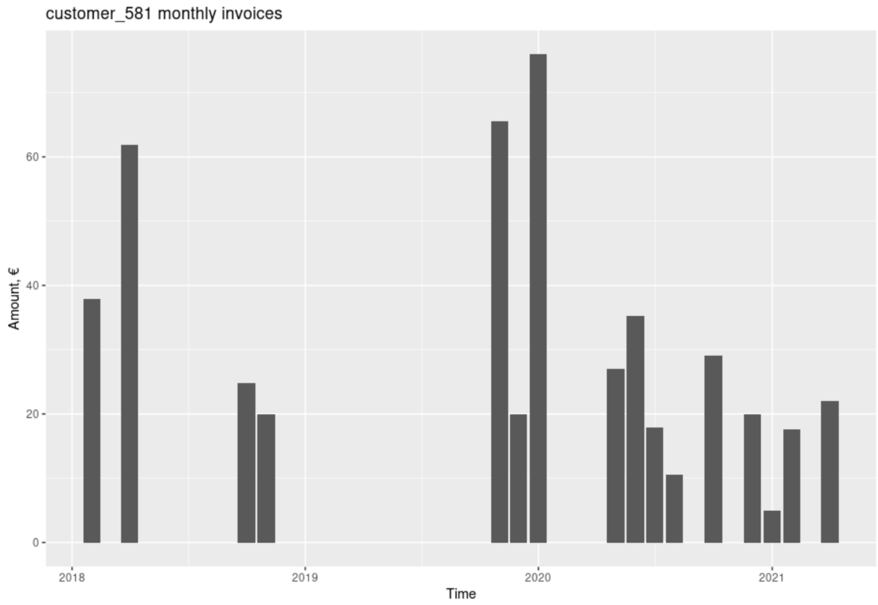

Figure 12: Spending history of customer 581.

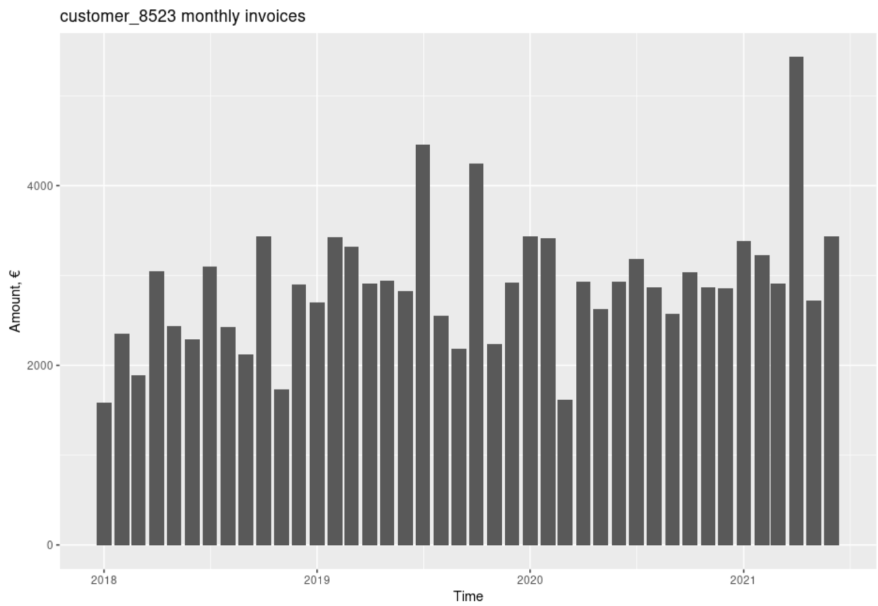

Figure 13: Spending history of customer 8523.

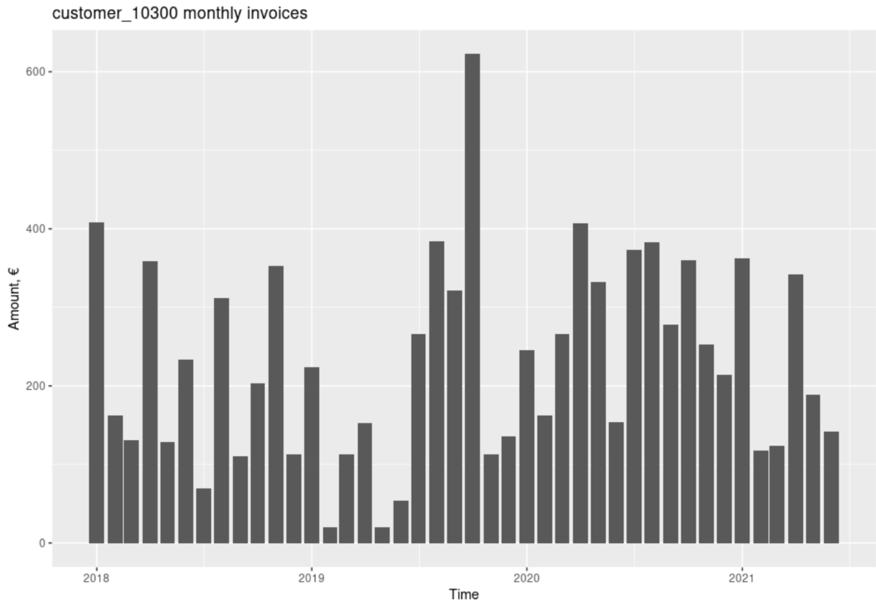

Figure 14: Spending history of customer 10300.

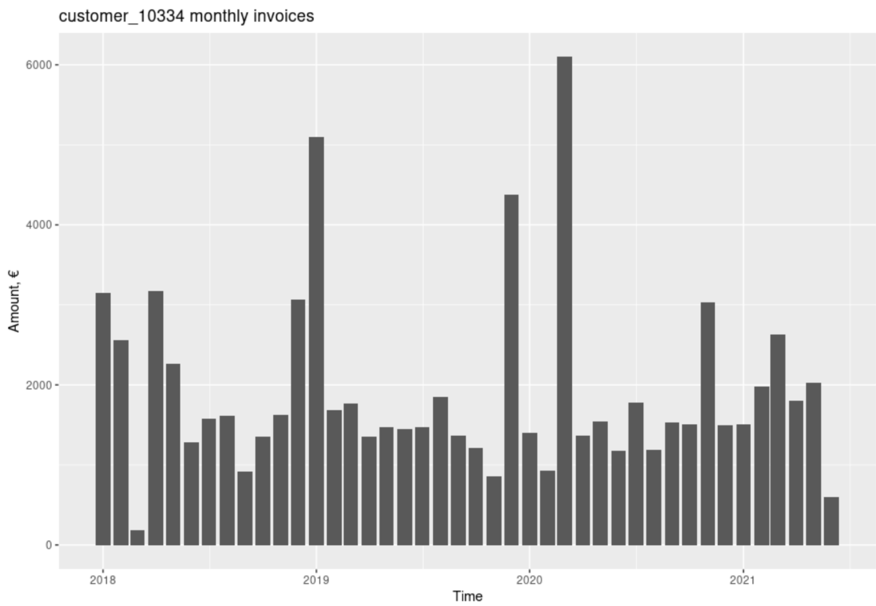

Figure 15: Spending history of customer 10334.

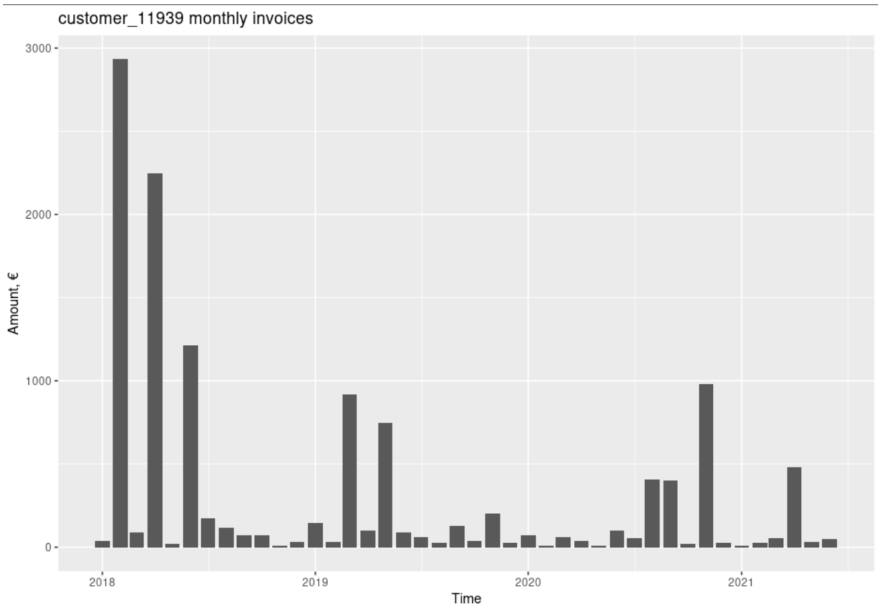

Figure 16: Spending history of customer 11939.

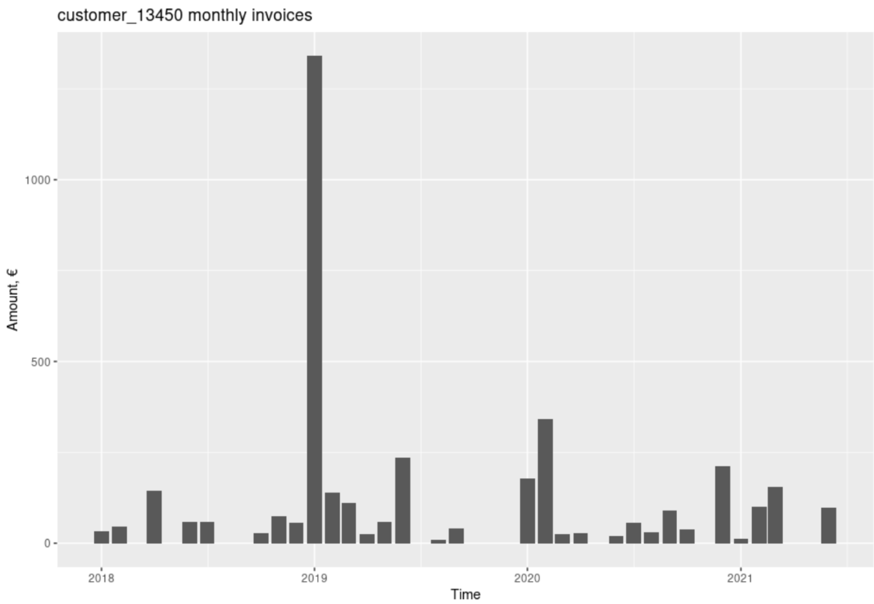

Figure 17: Spending history of customer 13450.

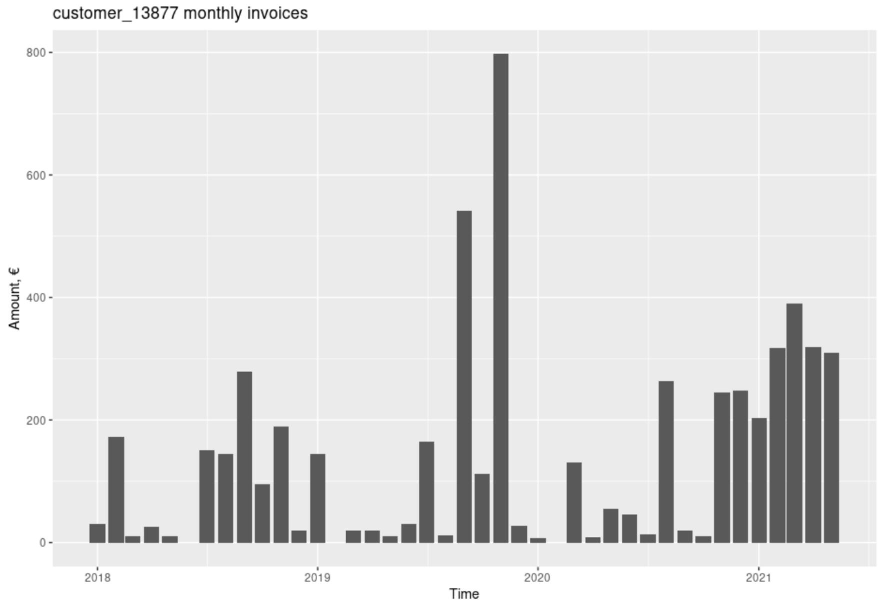

Figure 18: Spending history of customer 13877.

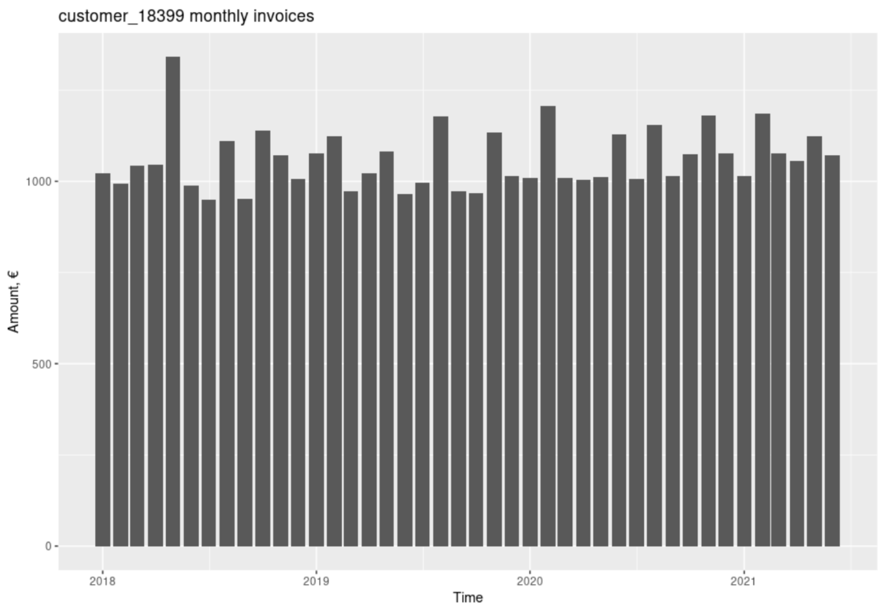

Figure 19: Spending history of customer 18399.

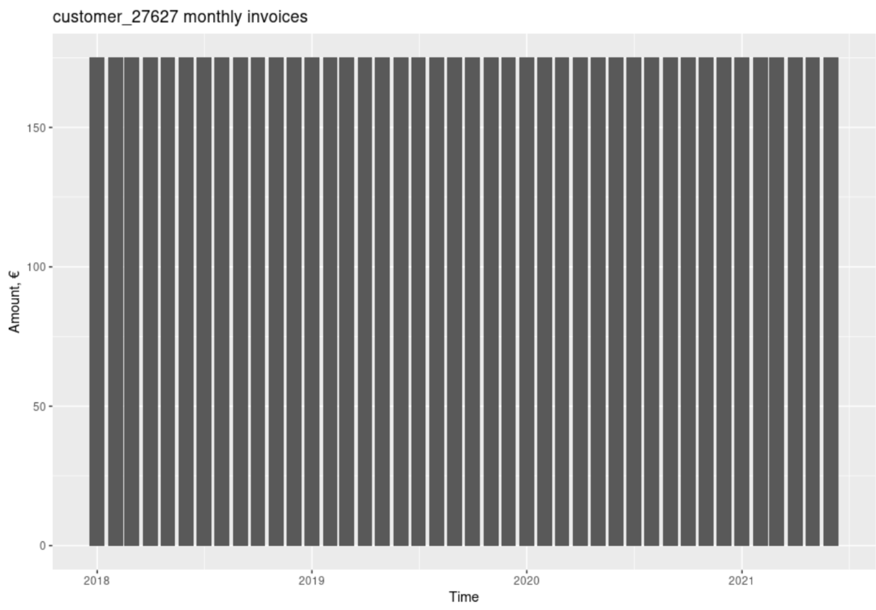

Figure 20: Spending history of customer 27627.

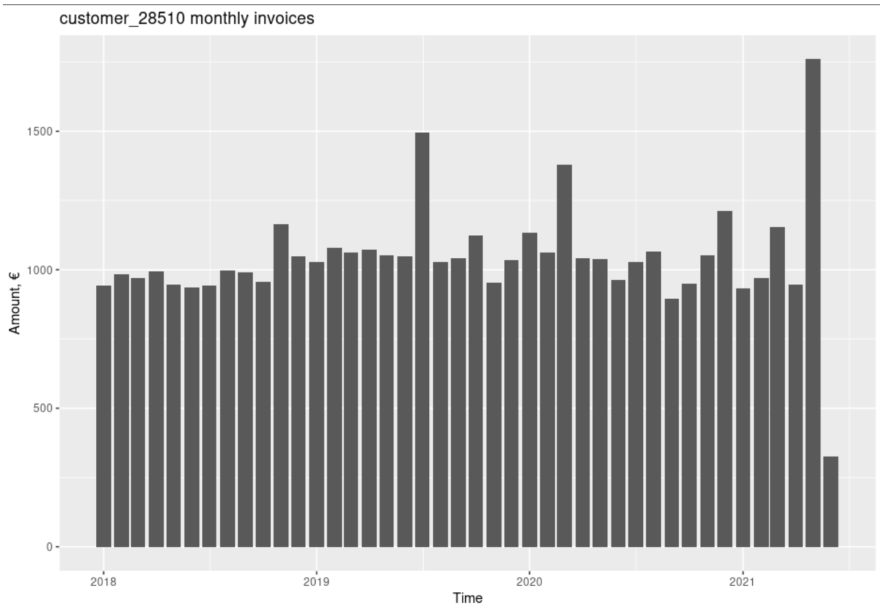

Figure 21: Spending history of customer 28510.

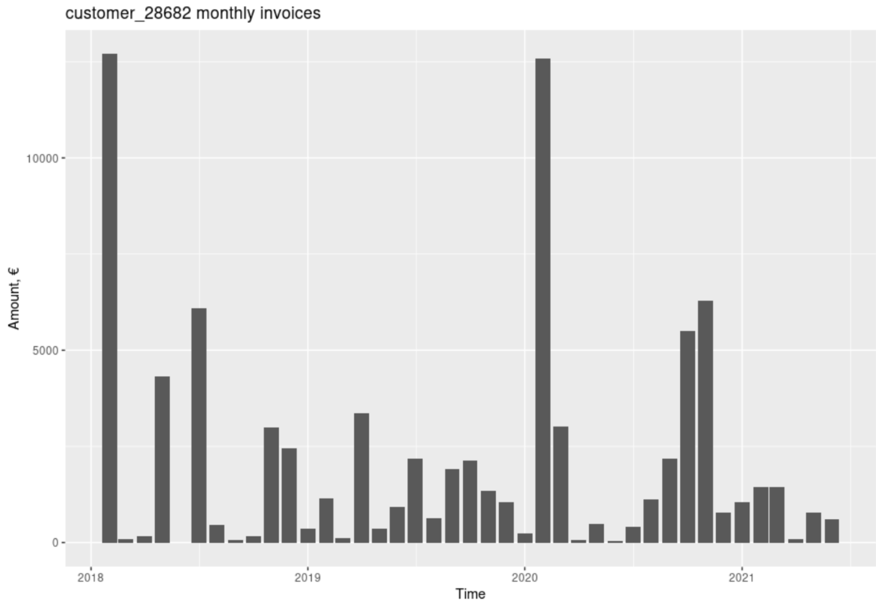

Figure 22: Spending history of customer 28682.

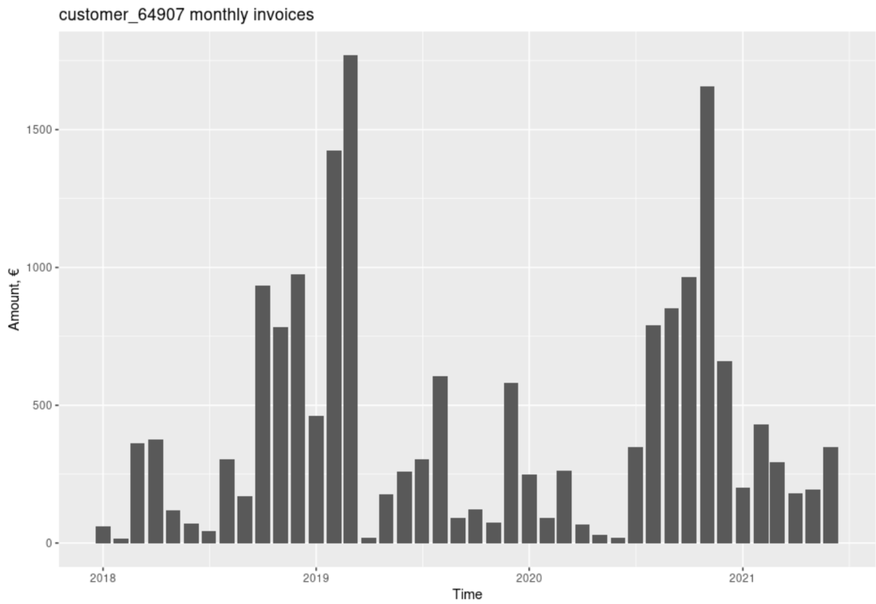

Figure 23: Spending history of customer 64907.

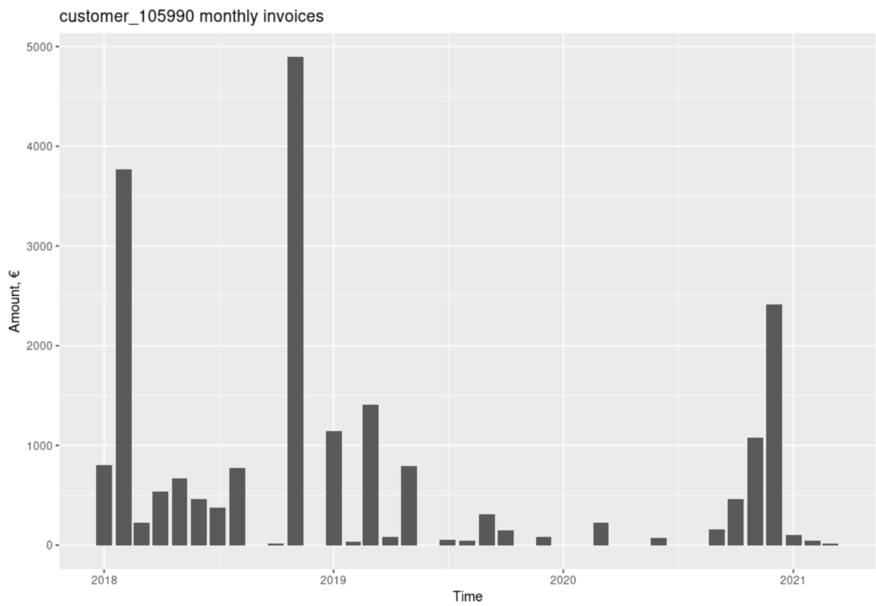

Figure 24: Spending history of customer 105990.