Ministry of Education and Science of the Republic of Kazakhstan

Astana IT University

FINAL REPORT

On the Completion of the Industrial Internship Program

Place of internship: TOO «AD Accounting Almaty»

Student: Baimuratova Kamila Muratkhanovna

Bar code: 230170

Practice period: 09.06-05.07

Educational Program: Big Data Analysis

Group: BDA-2306

Head of industrial internship: Ainur Mukashova

List of Contents

Calendar Plan

Приложение 2

КАЛЕНДАРНЫЙ ПЛАН-ГРАФИК ПРОХОЖДЕНИЯ ПРАКТИКИ ОТ ПРЕДПРИЯТИЯ

(составлять понедельно)

Баймуратова Камила Муратхановна, студента 2 курса, образовательной программы «Big Data Analysis», период практики с «9» июня 2025 г. по «5» июля 2025 г.

№ нед.	Наименование работ	Сроки исполнения	Название подразделения или рабочего места	Отметка об исполнении
1	2	3	4	5
1 нед.	Ознакомление с особенностями рабочей системы, онбординг	09.06-13.06	ТОО "AD Accounting Almaty"	выполнено
2 нед.	Изучение функционала платформ для работы	16.06-20.06	ТОО "AD Accounting Almaty"	выполнено
3 нед	Работа с базой данных: постановка задач и целей, чистка, подготовка и процесс обработки	23.06-27.06	ТОО "AD Accounting Almaty"	выполнено
4 нед	Визуализация полученных результатов и написание отчетов	30.06-05.07	ТОО "AD Accounting Almaty"	выполнено

Руководитель практики от предприятия _____ *Турткараева Н.К*

Руководитель практики от AITU _____ *Мукашова А.О.*

Подпись обучающегося _____ *Баймуратова К.М.*

Internship Reference

ТОО "AD Accounting Almaty"

Характеристика с места прохождения практики

Студент Баймуратова Камила Муратхановна за период прохождения стажировки проявила себя как ответственный, инициативный и целеустремленный сотрудник. В установленные сроки успешно освоила основные направления работы, активно интересовалась новыми инструментами и подходами в сфере анализа данных.

Камила ответственно подходила к выполнению поставленных задач: демонстрировала логическое мышление, грамотно использовала полученные теоретические знания на практике. При выполнении заданий проявила внимательность к деталям, аккуратность в оформлении отчетности и уверенное владение инструментами визуализации и обработки данных.

В коллективе зарекомендовала себя как открытый и доброжелательный человек, легко шла на контакт, проявляла уважение к коллегам. Все задачи, поставленные в рамках стажировки, были выполнены в полном объеме и в установленные сроки. По результатам прохождения стажировки может быть рекомендована к дальнейшему профессиональному росту в области анализа данных.

Рекомендуемая оценка: 100

Туркараева Н.К.

ТОО AD Accounting Almaty, директор на основании Устава

AD Accounting Almaty

М.П.

«27» Июня 2025 г.

Introduction

The internship at TOO “AD Accounting Almaty”, a professional outsourcing company providing accounting and financial services in Kazakhstan, offered an invaluable opportunity to apply data analytics skills in a real-world business environment. TOO “AD Accounting Almaty” serves a diverse portfolio of clients by offering bookkeeping, payroll, tax reporting, and financial consulting services, making it a dynamic setting for understanding data-driven operations in the financial sector.

During my internship, I applied the knowledge I gained during my studies, and if there were unclear tasks, I learned to solve them as problems arose. As a Data Analytics Intern, I was integrated into the analytics and reporting processes supporting the company's internal financial workflows and client reporting systems. My primary responsibilities included cleaning and structuring accounting data, developing automated reports in Microsoft Excel and Power BI, and assisting with the improvement of financial data pipelines to enhance reporting efficiency and accuracy. I was given work with real data from the 1C system, such as the balance sheet, the work was always connected with numbers.

Throughout the internship, I worked with tools such as Microsoft Excel ,Google Tools (Sheet, Collab, Drive, Docs), Power Query, and Power BI, applying techniques in ETL, data cleaning, and dashboard creation.

This practical experience reinforced key concepts from my Data Analytics coursework, including structured data manipulation, visualization, and the importance of data quality in financial reporting.

Moreover, it demonstrated how data analytics contributes to operational efficiency, client satisfaction, and compliance in the accounting domain.

Main part

1. Onboarding

The first weeks of my internship were relatively smooth, primarily focused on familiarizing myself with the new work system and environment. One of the most significant aspects of this period was becoming acquainted with the 1C system. 1C:Enterprise is described as "a universal cloud and on-premise system of programs for automating a company's financial and wider operational activities. 1C:Enterprise has the breadth of capability to address the diverse needs of today's business" (1C:Enterprise, n.d.).

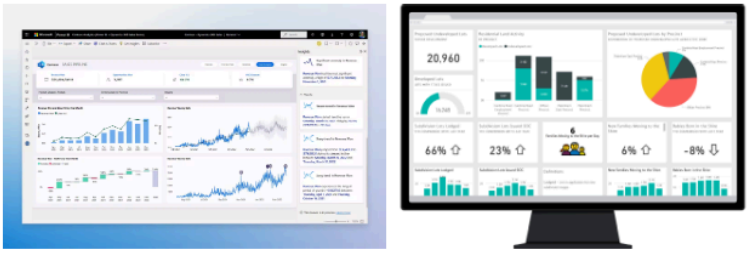
Although the system was not entirely new to me, I gained valuable insights. I learned how to extract and process data from reports, manage documents and orders, and understand how the system integrates with other programs for accounting and data analysis. The 1C system provides powerful tools for both accounting and analytics, enabling the rapid generation of reports and the performance of data analysis. I became familiar with how to configure filters and automate processes, which significantly enhances work efficiency.

Subsequently, I was assigned the task of preparing a report on key concepts, such as Power BI, Power Query, Statistics, and the differences between working in Google Sheets and Excel. This assignment was intended to help me refresh my knowledge of these tools and ease my transition into the workflow.

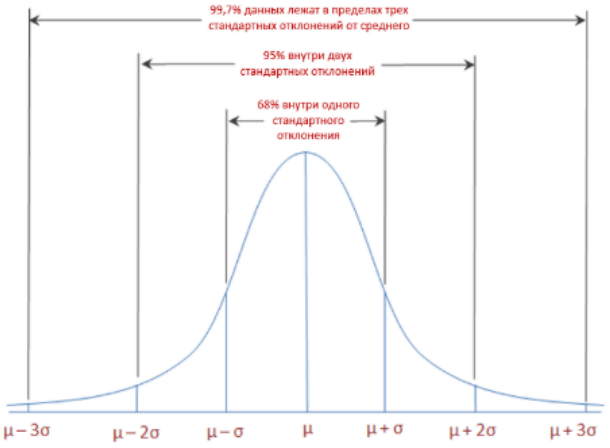
Отчет по основным терминам в сфере Дата Анализа

Стажер: Баймуратова Камила

Power BI - это мощный инструмент бизнес-анализа от Microsoft, который используется для визуализации и анализа данных. Он позволяет подключаться к различным источникам данных и создавать интерактивные отчеты и панели мониторинга (Microsoft, 2021). Power BI известен своими возможностями по визуализации данных и простоте использования.



Power Query - это технология, позволяющая пользователям преобразовывать и соединять данные из разных источников перед их загрузкой в Power BI или Excel. Power Query упрощает процесс подготовки данных, предлагая удобный графический интерфейс для их очистки и трансформации без необходимости писать код (Microsoft, 2021).



Различия между Google Sheets и Excel

Google Sheets позволяет нескольким пользователям одновременно работать с документом в реальном времени, что делает его удобным для совместной работы. В отличие от этого, Excel в основном работает как настольное приложение (но есть и онлайн-версии с ограниченными функциями совместной работы) (Microsoft, 2021).

Excel предлагает более мощные инструменты для анализа данных, такие как сводные таблицы и более сложные функции, в то время как Google Sheets лучше подходит для базовых задач (Microsoft, 2021).

Figure 1 and 2. Quick report on key terms in date analysis

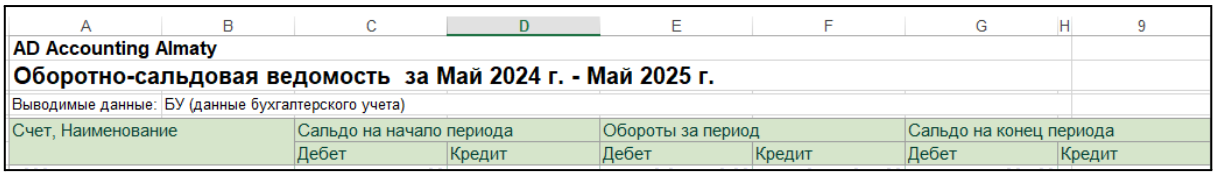
An essential component of the internship experience was the onboarding process, which involved getting to know the team and the individuals I would be working with over the course of the next month. The adaptation process was successful, and I quickly established rapport with my colleagues and supervisor. They were supportive, providing explanations for any unclear points, and ensured that I was not overwhelmed with tasks, which allowed for a smooth integration into the work environment.

2. ETL

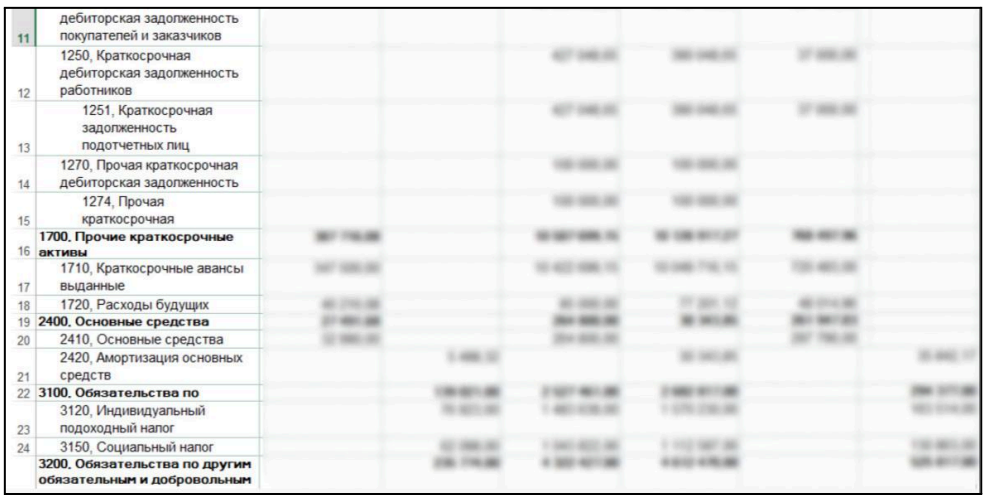
My project involved analyzing the general and client turnover-balance sheet for the past year.

E - Extract: To start, the first step was to extract all the relevant data from the 1C system. This part of the process was relatively straightforward. I simply needed to navigate to the required section in the system and click the "export to Excel" option. The file was initially saved on my device in the older xls format. Afterward, I reformatted the file into the more modern xlsx version to ensure better compatibility with my analysis tools.

T - Transform: The next step, transformation, proved to be a bit more challenging. The main issue was the poor transfer of data into the table. Some cells had been incorrectly split, while others were merged, which caused a misalignment in the data. Additionally, there were unnecessary rows and columns inserted between the actual data, making it difficult to properly read the columns, headers, and the overall structure of the information. This created several obstacles that required extra time and effort to fix in order to ensure the data was in a usable format for further analysis.



A	B	C	D	E	F	G	H	9
AD Accounting Almaty								
Оборотно-сальдовая ведомость за Май 2024 г. - Май 2025 г.								
Выводимые данные:	БУ (данные бухгалтерского учета)							
Счет, Наименование	Сальдо на начало периода		Обороты за период		Сальдо на конец периода			
	Дебет	Кредит	Дебет	Кредит	Дебет	Кредит		



11	дебиторская задолженность покупателей и заказчиков						
12	1250, Краткосрочная дебиторская задолженность работников						
13	1251, Краткосрочная задолженность подотчетных лиц						
14	1270, Прочая краткосрочная дебиторская задолженность						
15	1274, Прочая краткосрочная						
16	**1700, Прочие краткосрочные активы**						
17	1710, Краткосрочные авансы выданные						
18	1720, Расходы будущих						
19	**2400, Основные средства**						
20	2410, Основные средства						
21	2420, Амортизация основных средств						
22	**3100, Обязательства по**						
23	3120, Индивидуальный подоходный налог						
24	3150, Социальный налог						
	3200, Обязательства по другим обязательным и добровольным						

Figures 3 and 4 present the balance sheet for the year of the company "AD Accounting Almaty" (the figures are classified as they represent real company data, for the purpose of maintaining confidentiality).

First, using Python and the pandas library, I cleaned the data by removing empty rows and columns, converted specific columns into numerical format, and saved the result into an Excel file. Later, I manually adjusted the alignments and split or merged cells where necessary to ensure that the data was properly structured. This step was important because while Python and pandas helped automate the initial cleaning, some adjustments required manual intervention to ensure the file was perfectly formatted for analysis. The combination of automation and manual corrections allowed me to prepare the dataset in a more efficient and accurate way for further processing.

```
[ ] import pandas as pd

    from google.colab import files
    uploaded = files.upload()
```

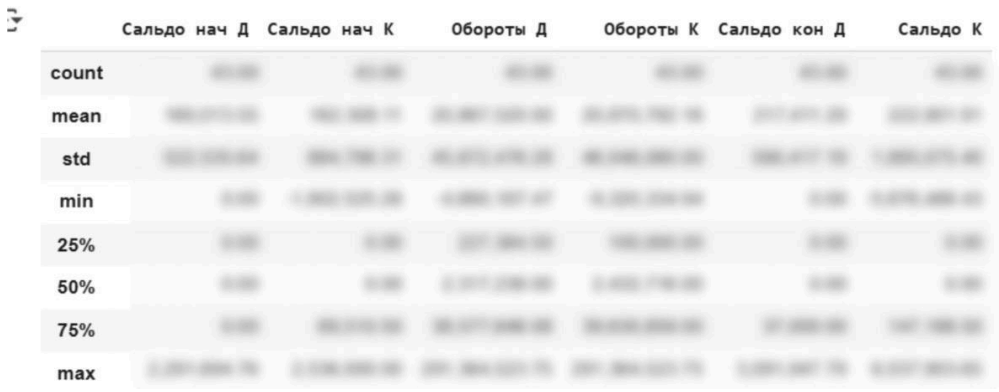
```
Выбрать файлы  Файл не выбран    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving dolg1.xlsx to dolg1.xlsx
```

```
[ ] data = pd.read_excel('dolg1.xlsx', sheet_name=None)
```

```
df = data[list(data.keys())[0]]

df = df.dropna(how='all', axis=0)
df = df.dropna(how='all', axis=1)

df['Сальдо на начало периода'] = df['Сальдо на начало периода'].apply(pd.to_numeric, errors='coerce')
df['Обороты за период'] = df['Обороты за период'].apply(pd.to_numeric, errors='coerce')
df['Сальдо на конец периода'] = df['Сальдо на конец периода'].apply(pd.to_numeric, errors='coerce')

output_file = '/mnt/data/formatted_table.xlsx'
df.to_excel(output_file, index=False)


output_file
```

figure 5. data cleaning

L - Load: Next, I upload the cleaned file into Google Colab and begin working on the analysis. I load the data into the environment and start by re-checking the data types, ensuring there are no missing values, and performing an initial statistical overview. This step is essential to verify the integrity of the data before diving deeper into the analysis. I carefully inspect each column to make sure that the values are properly formatted and that there are no inconsistencies. This preparation ensures that the analysis is based on accurate and complete data, minimizing the risk of errors during the more advanced stages of the project.



	Сальдо нач Д	Сальдо нач К	Обороты Д	Обороты К	Сальдо кон Д	Сальдо К
count	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
mean	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
std	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
min	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25%	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
50%	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
75%	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
max	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

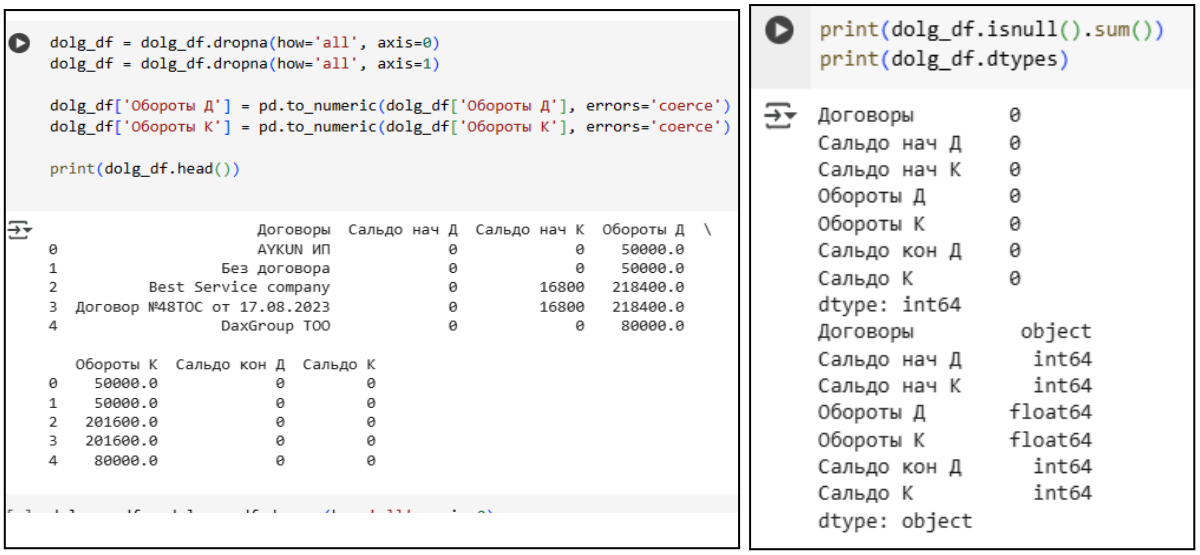
figure 6. statistical metrics (the figures are classified as they represent real company data, for the purpose of maintaining confidentiality).

Next, I tried to forecast the predicted debit turnover for new data, but since I only had access to data for one year, the MAE was too large. However, the forecast was positive, even considering the biggest negative value and the error.

After that, I started working with the data on debts from clients. Our clients are individual entrepreneurs (IE/ИП) and limited liability partnerships (LLP/ТОО), who come to us to avoid hiring an accountant themselves. When debts accumulate on their side, our credit increases.

First, I link my Google Drive to Google Colab in order to upload the files I need. Then, I begin the data cleaning process by removing empty rows or columns, filling in missing values, and, of course, performing thorough checks to ensure the data is accurate and ready for analysis.

This process of cleaning and preparing data is crucial to ensure the accuracy and reliability of the analysis, as errors in the data could lead to misleading conclusions or incorrect financial decisions.



```
dolg_df = dolg_df.dropna(how='all', axis=0)
dolg_df = dolg_df.dropna(how='all', axis=1)

dolg_df['Обороты Д'] = pd.to_numeric(dolg_df['Обороты Д'], errors='coerce')
dolg_df['Обороты К'] = pd.to_numeric(dolg_df['Обороты К'], errors='coerce')

print(dolg_df.head())
```

```
                      Договоры  Сальдо нач Д  Сальдо нач К  Обороты Д  \
0                    AYKUN ИП             0             0    50000.0
1                Без договора             0             0    50000.0
2          Best Service company             0         16800   218400.0
3  Договор №48ТОС от 17.08.2023             0         16800   218400.0
4                 DaxGroup ТОО             0             0    80000.0

   Обороты К  Сальдо кон Д  Сальдо К
0    50000.0             0         0
1    50000.0             0         0
2   201600.0             0         0
3   201600.0             0         0
4    80000.0             0         0
```

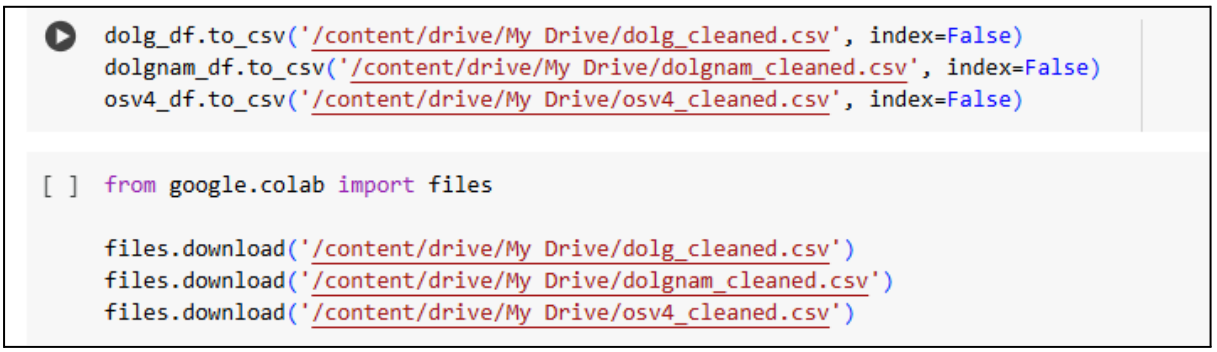
```
print(dolg_df.isnull().sum())
print(dolg_df.dtypes)
```

```
Договоры        0
Сальдо нач Д    0
Сальдо нач К    0
Обороты Д       0
Обороты К       0
Сальдо кон Д    0
Сальдо К        0
dtype: int64
Договоры         object
Сальдо нач Д      int64
Сальдо нач К      int64
Обороты Д       float64
Обороты К       float64
Сальдо кон Д      int64
Сальдо К          int64
dtype: object
```

figure 7 and 8. code snippets

At the same time, I was cleaning an identical file, but this one contained data on receivables. After completing the cleaning process, I saved the final CSV files and moved on to continue my work.

Once the data was cleaned and saved, I felt confident that it was ready for further analysis and reporting.



```
dolg_df.to_csv('/content/drive/My Drive/dolg_cleaned.csv', index=False)
dolgnam_df.to_csv('/content/drive/My Drive/dolgnam_cleaned.csv', index=False)
osv4_df.to_csv('/content/drive/My Drive/osv4_cleaned.csv', index=False)
```

```
[ ] from google.colab import files

    files.download('/content/drive/My Drive/dolg_cleaned.csv')
    files.download('/content/drive/My Drive/dolgnam_cleaned.csv')
    files.download('/content/drive/My Drive/osv4_cleaned.csv')
```

figure 9. code snippets

3. Dashboard building

I was then tasked with building a dashboard for our company's debts in Power BI. However, since I had never worked with this tool before, I was advised to first practice, study the functions, and understand how the buttons and features work. As a training assignment, I created my first dashboard in Power BI using the sample dataset sales_data.csv, which contained fictional sales data for an online store. My main goal was to learn the basic functions of Power BI: loading data, creating visualizations, setting up metrics, and producing an analytical report in the form of a dashboard.

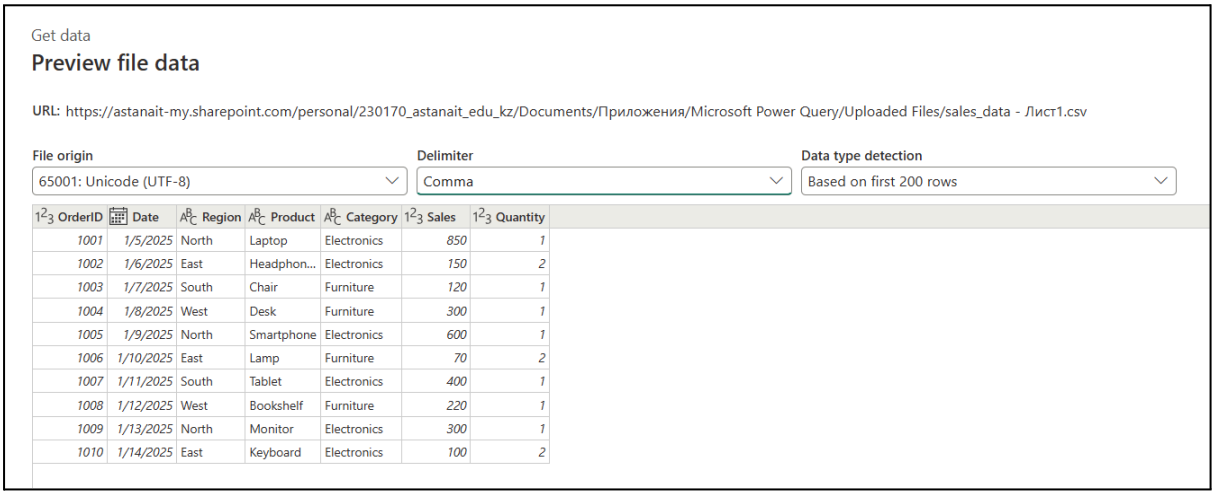
During the process of creating my first dashboard in Power BI, I used several sources that helped me learn the basics of the tool and how to create visualizations and reports. In this process, I relied on Microsoft's official documentation, video tutorials, and practical books, which allowed me to effectively study all the necessary features.

The primary source of information for learning Power BI was the official Microsoft documentation, which provided a comprehensive guide to both basic and advanced Power BI features, such as creating and configuring visualizations, loading data, creating reports, and dashboards (Microsoft, n.d.). This documentation helped me gain a deeper understanding of how to properly work with data and the capabilities Power BI offers for analysis.

To complement my theoretical knowledge, I also used video tutorials on the YouTube channel Kevin Sratvert. The channel provides numerous useful lessons on Power BI, covering both basic and more advanced aspects of working with the tool. These videos were an excellent complement to the articles and documents, allowing me to see how to apply the knowledge in practice and accelerating my learning process (Kevin Sratvert, 2023).

To start, I manually prepared the CSV file, including fields such as order number, date, region, product, category, sales amount, and quantity. I then added calculated fields: cost, profit, profit margin,

average order value, as well as year and month of the order, to enable further analysis by time and category.


Get data

Preview file data

URL: https://astanait-my.sharepoint.com/personal/230170_astanait_edu_kz/Documents/Приложения/Microsoft Power Query/Uploaded Files/sales_data - Лист1.csv

File origin: 65001: Unicode (UTF-8) | Delimiter: Comma | Data type detection: Based on first 200 rows

OrderID	Date	Region	Product	Category	Sales	Quantity
1001	1/5/2025	North	Laptop	Electronics	850	1
1002	1/6/2025	East	Headphon...	Electronics	150	2
1003	1/7/2025	South	Chair	Furniture	120	1
1004	1/8/2025	West	Desk	Furniture	300	1
1005	1/9/2025	North	Smartphone	Electronics	600	1
1006	1/10/2025	East	Lamp	Furniture	70	2
1007	1/11/2025	South	Tablet	Electronics	400	1
1008	1/12/2025	West	Bookshelf	Furniture	220	1
1009	1/13/2025	North	Monitor	Electronics	300	1
1010	1/14/2025	East	Keyboard	Electronics	100	2

figure 10. dataset for training

Once the file was prepared, I uploaded it to Power BI Web. I selected the option "Create" → "CSV (Preview)", which allowed me to quickly import the data and begin working with it. The next step was to create the visualizations. The dashboard included the following elements: a bar chart to display sales by region, a pie chart to analyze sales by product category, a line chart to analyze sales trends over time, and a card showing the total sales amount. I placed all the visual elements on a single page and grouped them under the title "Online store sales dashboard" for better report presentation.

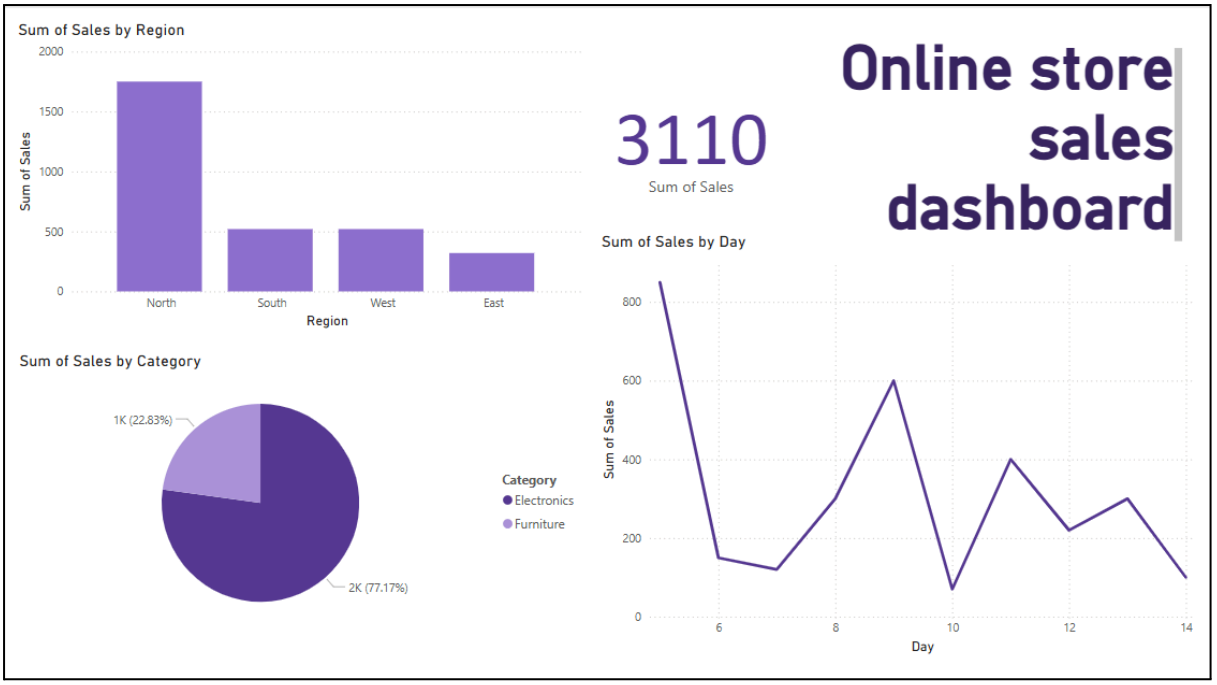


figure 11. dashboard created for training

For my first dashboard, the result turned out to be quite good. Although the visualizations and analysis are simple, they effectively show the main sales data, such as distribution by region, product category, and sales trends over time. All the visual elements are presented on one page, and the information is easy to understand. Of course, this is a basic report, but for a first dashboard, it's a pretty good start. I was able to master the basic functions of Power BI and create a clear and understandable report. In the future, more complex metrics and advanced visualizations can be added, but at this stage, I'm satisfied with the result.

This process helped me familiarize myself with the capabilities of Power BI, learn how to build simple reports, use different visualizations, and work with key metrics. Additionally, I understood the importance of properly structuring data for successful analysis.

After learning how to use Power BI, I began my task that was given by the company. During the course of the project, I used several tools to process and analyze the data. Microsoft Excel was applied to work with the initial accounting tables and to perform basic calculations of debit and credit turnovers. For more advanced data manipulation, I used Power Query, which allowed me to extract the necessary data,

clean it from empty or irrelevant values, convert data types, and group information by contracts for easier analysis. After preparing the dataset, I imported it into Power BI, where I created a visual dashboard that clearly represents the dynamics of client debt over the selected period. The source of all the information was the company's financial data, which included details on contracts, balances at the beginning and end of the period, as well as turnover values.

The initial stage involved extracting and processing the source table in Power Query. The following steps were performed:

1. Removal of empty rows and columns
2. Data type conversion
3. Grouping data by contracts
4. Calculating intermediate totals for debit and credit.

Three key visualizations were built using Power BI:

- Opening Balance (Credit): shows the starting debt of clients as of May 2024. The largest amounts were from IP MRES and the contract dated 27.09.2023.
- Closing Balance (Credit): shows the growth of debt by May 2025
- Debit and Credit Turnovers: provides a comparative analysis of incoming and outgoing amounts by contract.

figure 12. customer debt dashboard (the figures are classified as they represent real company data, for the purpose of maintaining confidentiality)

Tools and technologies

During the creation of the dashboard and throughout my work with data analysis, I utilized several tools and technologies that helped streamline the process and enhance the accuracy of my results.

I began by working with Microsoft Excel and Google Sheets, which allowed me to manually prepare and clean the dataset before importing it into more advanced tools. Both platforms are essential for data manipulation, such as organizing and calculating fields, handling large datasets, and performing basic analysis. Next, I worked with Power BI, which served as the primary tool for creating the dashboard. Power BI allowed me to visualize the data in a variety of formats, including bar charts, line graphs, and pie charts. It provided a powerful way to analyze and present data interactively and in a visually engaging manner. Additionally, I utilized Power Query within Power BI for transforming and cleaning the data before it was used in the final visualizations. Power Query enabled me to perform tasks such as data reshaping, merging tables, and applying calculated columns, which streamlined the process of data preparation. For advanced analysis, I used Jupyter Notebook and Google Colab, which are both excellent environments for working with Python-based data analysis libraries such as Pandas, NumPy, and Matplotlib. These tools provided me with the flexibility to process and analyze data programmatically, offering greater control over the analysis and enabling more complex calculations.

Technical Contributions

In this project, my primary technical contributions involved preparing, cleaning, and transforming data for visualization and analysis. Initially, I used Excel and Google Sheets to organize the dataset and clean it by removing duplicates, handling missing values, and calculating new fields such as cost, profit, profit margin, and average order value. This data preparation ensured that the dataset was in an appropriate format for further processing and visualization.

After preparing the data, I utilized Power BI to create various visualizations that effectively conveyed the data insights. I developed a set of visual elements, including bar charts for sales by region, pie charts for product category distribution, and line charts to display sales trends over time. Additionally, I used Power Query within Power BI to perform more advanced data transformations, such as merging datasets, applying filters, and reshaping data for more accurate analysis and reporting.

Furthermore, I employed Jupyter Notebook and Google Colab to perform additional data analysis using Python. These platforms allowed me to apply libraries such as Pandas, NumPy, and Matplotlib for advanced statistical analysis and data manipulation. This enabled me to dive deeper into the dataset, extract valuable insights, and refine the visualizations I created in Power BI, ensuring a comprehensive and accurate final dashboard.

Challenges and solutions

One of the main challenges I faced during my internship was the complete lack of prior experience in the field. As I had never worked in the industry before, it took me some time to fully understand the tasks and responsibilities assigned to me. Initially, this slowed my progress as I had to learn everything from scratch. However, this challenge became an opportunity for growth. Step by step, I gained the necessary knowledge and skills, building a solid foundation that would serve me well in the future. I sought help from my colleagues when needed and made sure to ask questions to clarify doubts, which eventually helped me catch up and work more efficiently.

Another significant challenge was the industry I was interning in — accounting. With no background in this area, I had to familiarize myself with financial terminology, concepts, and systems specific to the field. This was a steep learning curve, and I dedicated time to independent study, reading relevant materials, and consulting with my colleagues to understand the context better. Over time, I became more comfortable with the concepts and tools used in accounting, and this allowed me to perform my tasks more effectively. Additionally, I improved my understanding of how financial data analysis plays a crucial role in decision-making and business operations.

A more personal challenge I faced was the age gap between me and my colleagues. As the youngest member of the team, I initially felt apprehensive about asking questions or appearing inexperienced. I feared that my lack of experience might be seen as a weakness. However, with time and support from my team, I gained more confidence and learned how to communicate and collaborate professionally. The team was always supportive, and their encouragement helped me overcome my initial fears, ultimately allowing me to build stronger working relationships with my colleagues.

Reflection and conclusion

Despite these challenges, I am grateful for the opportunity to intern at the company. The support from my supervisors and colleagues was invaluable, and it played a key role in my growth during this internship. I particularly want to express my appreciation to my supervisor, whose guidance and expertise helped me navigate difficult situations and stay focused on my learning goals.

Overall, the internship was an incredibly valuable experience. While there were challenges along the way, I successfully completed my tasks, and my supervisors were satisfied with my performance. I gained practical experience with tools like Jupyter Notebook, Power BI, Excel, and Power Query, which enhanced my technical skills. This internship not only improved my ability to analyze data but also gave me a deeper understanding of how data analysis is applied in real-world business environments. Most importantly, it has made me feel more confident and prepared to work in a professional IT setting and continue developing my skills as a data analyst.

References

1C Company. (n.d.). *What is 1C:Enterprise?* 1C:Developer Network. https://1c-dn.com/1c_enterprise/what_is_1c_enterprise/

Kevin Stratvert. (2023, February). *Power BI Tutorial for Beginners* [Video]. YouTube. https://www.youtube.com/watch?v=NNSHu0rkew8

Microsoft. (n.d.). *Power BI documentation*. Microsoft Learn. https://docs.microsoft.com/en-us/power-bi/