Problem:

For the past few years, Yellevate has been struggling with client disputes. They have defined disputes as clients expressing dissatisfaction with the company's services and refusing to pay for them. This has been a huge financial burden for the company.

Statistically, nearly 20% of the disputes raised against the company resulted in a payment opt-out. This has led to an approximate 5% annual loss of revenue (in USD). And now, the management has decided to approach the data analyst team to help resolve the issue.

The team was tasked to help identify the causes of the disputes from the data the company has collected.They are also tasked to provide in-depth analysis and actionable strategies to help resolve the problem.

The Executives at the company have decided that the following information should be obtained to identify the circumstances around the dispute problem:

1. The processing time in which invoices are settled (average # of days rounded to a whole number).
2. The processing time for the company to settle disputes (average # of days rounded to a whole number).
3. Percentage of disputes received by the company that were lost (within two decimal places).
4. Percentage of revenue lost from disputes (within two decimal places).
5. The country where the company reached the highest losses from lost disputes (in USD).

Methodology:

The purpose of this report is to gain insights with client disputes that have been causing a huge financial burden for the company, Yellevate.

The data analyst team has been tasked to help identify the causes of the disputes from the data the company has collected, and in turn provide necessary solutions and recommendations.

The company has provided this data and the data dictionary for the team to fully grasp the information within the dataset and the necessary terms to describe each column and its purpose.

With the data provided, we noticed that the inputs are a mix of discrete quantitative data and nominal data. With this in mind, we have decided to check the data dictionary to validate which column we should focus on to get the answers to the data analysis goals.

We have decided to analyze and process the provided inputs by running them in two different programs (Microsoft Excel & pgAdmin4). In pgAdmin4, we have imported the dataset and checked if there is erroneous information or records that will affect the overall quality of data.

After running and cleaning the data, we have decided to import the results in Microsoft Excel and create the necessary tables and charts in order to distinguish and to make a conclusion that will satisfy the data analysis goals.

Cleaning and processing the data:

With the provided data dictionary, we noticed the columns 'disputed' and 'dispute_lost' both have values of 0 and 1.

In the column 'disputed', it is stated that the value of '1' signifies that the customer has disputed the invoice, while '0' means they did not.

In the column 'dispute_lost', it is stated that the value of '1' means Yellevate lost the dispute and the dispute was resolved in favor of the customer, and the customer does not have to pay the invoice. While the value of '0' means, the customer did not win the dispute, and they are legally required to pay the full invoice amount. The value of '0' also means that Yellevate won the dispute, or there was no dispute in the first place.

A. SQL

- In pgAdmin4, we have imported the data and created a table in order for us to check the values inside it. Here is the query/codes we have used:

```
CREATE TABLE IF NOT EXISTS yellevate_invoices (
        country VARCHAR,
        customer_ID VARCHAR,
        invoice_number NUMERIC,
        invoice_date DATE,
        due_date DATE,
        invoice_amount NUMERIC,
        disputed NUMERIC,
        dispute_lost NUMERIC,
        settled_date DATE,
        days_to_settle INTEGER,
        days_late INTEGER
);
```

- We checked the data for duplicates using the aggregate function COUNT, and the HAVING clause to return duplicate values.

```
SELECT customer_id, invoice_number, COUNT(*) FROM yellevate_invoices
GROUP BY customer_id, invoice_number
```

```
HAVING COUNT (*) > 1;
```

- For data analysis goal 1, we have used the aggregate function ROUND and the GROUP BY clause, to compute for the processing time in which the invoices are settled.

```
SELECT country, ROUND(AVG(days_to_settle)) AS average_days_to_settle FROM
yellevate_invoices
GROUP BY country;

SELECT ROUND(AVG(days_to_settle)) AS days_settled
 FROM yellevate_invoices;
```

- For data analysis goal 2, we have used the aggregate functions ROUND, AVERAGE and the WHERE clause to compute for the processing time for the company to settle disputes

```
SELECT ROUND (AVG(days_to_settle)) AS days_to_settle
FROM yellevate_invoices
WHERE disputed = 1;

SELECT country, ROUND(AVG(days_to_settle)) AS settle_dispute_average_day
FROM yellevate_invoices
WHERE disputed = 1
GROUP BY country;
```

- For data analysis goal 3, we have used the aggregate functions ROUND, AVERAGE and the GROUP BY, HAVING, ORDER BY clauses to compute and filter out the percentage of disputes received by the company that were lost

```
SELECT country, ROUND (AVG(dispute_lost) * 100, 2) AS percentage_lost_disputes
FROM yellevate_invoices
GROUP BY country, disputed
HAVING disputed = 1
ORDER BY percentage_lost_disputes DESC;
```

```
SELECT ROUND (AVG(dispute_lost) * 100, 2) AS percentage_lost_disputes
FROM yellevate_invoices
WHERE disputed = 1;
```

- For data analysis goal 4, we have used the aggregate functions ROUND, SUM and the GROUP BY and HAVING clauses to compute for the percentage of revenue lost from disputes

```
SELECT country, ROUND(SUM(invoice_amount) / (SELECT SUM(invoice_amount)
FROM yellevate_invoices) * 100, 2) FROM yellevate_invoices
GROUP BY country, dispute_lost
HAVING dispute_lost = 1;
```

```
SELECT ROUND(SUM(invoice_amount) / (SELECT SUM(invoice_amount)
FROM yellevate_invoices) * 100, 2) FROM yellevate_invoices
GROUP BY dispute_lost
HAVING dispute_lost = 1;
```

- For data analysis goal 5, we have used the aggregate function SUM, the GROUP BY, HAVING and ORDER BY to compute and filter out the country where the company reached the highest losses from disputes.

```
SELECT country, SUM(invoice_amount) AS invoice_amount_lost_in_disputes
FROM yellevate_invoices
GROUP BY country, dispute_lost
HAVING dispute_lost > 0
ORDER BY SUM(invoice_amount) DESC;
```

B. Excel

- In Excel, we have imported the processed data from the now cleaned data from pgAdmin4 and made the pivot tables and necessary visualizations using charts.
- For data analysis goal 1, we have computed the processing time in which invoices are settled by averaging the number of ‘days_settled’. The results yielded to 26 days

Country	Average days settled
China	23
Spain	25

France	28
United States	28
Russia	29
Grand Total	26

- For data analysis goal 2, we have computed the processing time for the company to settle the disputes by averaging the number of ‘days_settled’ with a filter specifically pertaining only on the ‘disputed’. The results yielded to 36 days.

disputed	1

Country	Average days settled
France	34
China	34
Spain	37
Russia	38
United States	41
Grand Total	36

- For data analysis goal 3, we have computed the percentage of disputes received by the company that were lost by tallying the total count of ‘dispute_lost’ and placing ‘disputed’ as the filter and adjusting it to ‘1’, meaning we are only computing for the disputed value and then adding again ‘dispute_lost’ in the column in order to compute the percentage. The results yielded to 17.69%

disputed	1

Count of Lost Disputes	Column Labels		
Country	Not Disputed	Disputed	Grand Total
China	91.80%	8.20%	100.00%
France	65.77%	34.23%	100.00%
Russia	91.28%	8.72%	100.00%
Spain	93.22%	6.78%	100.00%
United States	96.25%	3.75%	100.00%
Grand Total	82.31%	17.69%	100.00%

- For data analysis goal 4, we have computed the percentage of revenue lost from disputes by tallying the sum of 'invoice_amount' and placing 'disputed' as filter and adjusting it to 'all', meaning we are computing for the value of all the disputes then adding again 'dispute_lost' in the column, then lastly, we have adjusted the value field settings of the 'sum of invoice_amount' to show the value as '% of row total' to show the percentage values. The results yielded to 4.67%

disputed	(All)

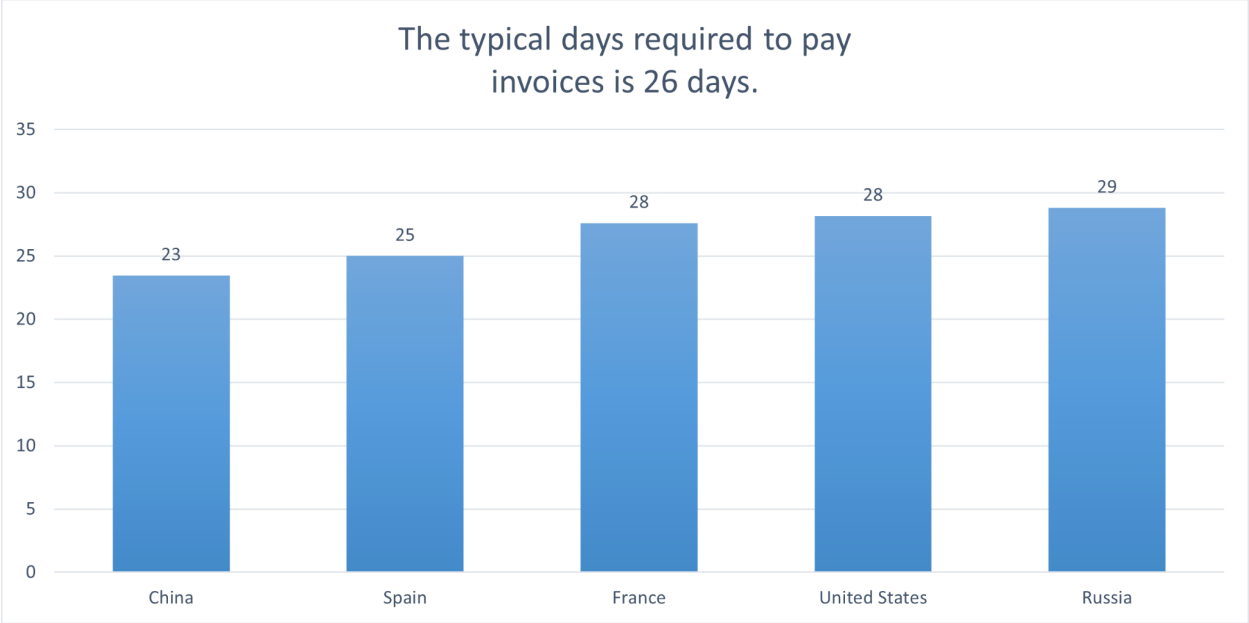
Sum of invoice_amount	Column Labels		
Country	Not Disputed	Disputed	Grand Total
Spain	98.96%	1.04%	100.00%
United States	99.16%	0.84%	100.00%
China	98.94%	1.06%	100.00%
Russia	96.68%	3.32%	100.00%
France	86.65%	13.35%	100.00%
Grand Total	95.33%	4.67%	100.00%

- For data analysis goal 5, we have computed for the country where the company reached the highest losses from disputes by tallying the sum of ‘invoice_amount’ and placing ‘disputed’ as filter and adjusting it to ‘all’, meaning we are computing for the value of all the disputes then adding again ‘dispute_lost’ in the column in order to compute the percentage. The results yielded to $526,264.

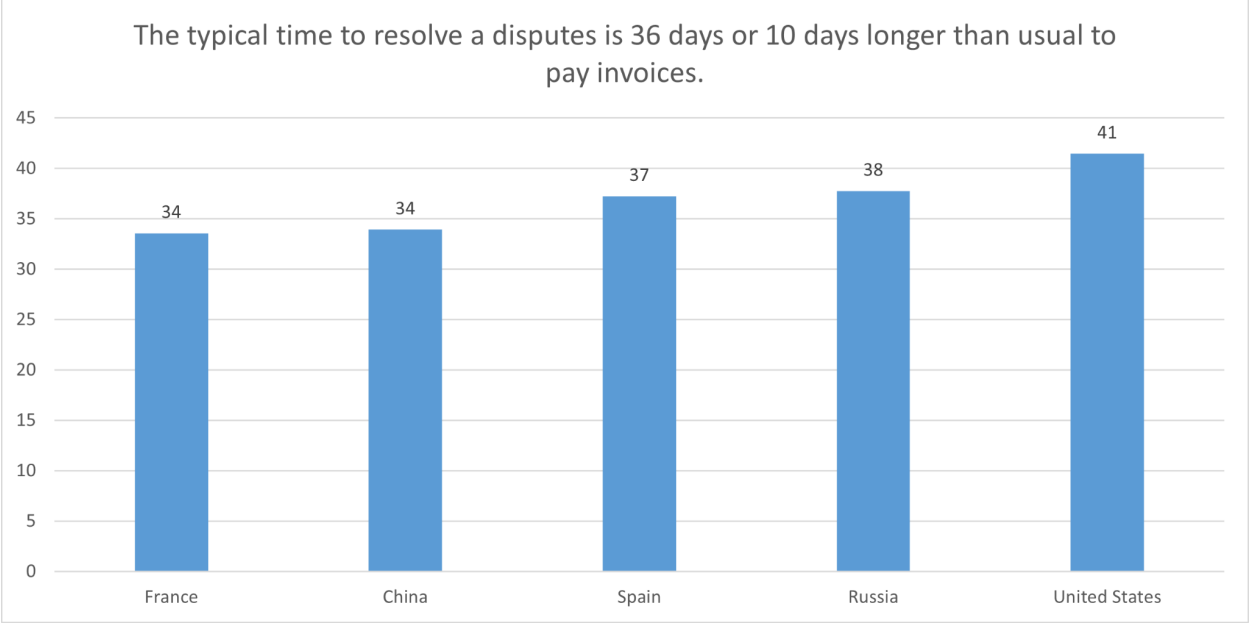
disputed	(All)

Sum of invoice_amount	Column Labels		
Country	Not Disputed	Disputed	Grand Total
Spain	$1,617,802	$17,046	$1,634,848
United States	$2,715,141	$22,936	$2,738,077
China	$3,962,266	$42,630	$4,004,896
Russia	$2,368,915	$81,291	$2,450,206
France	$3,416,027	$526,264	$3,942,291
Grand Total	$14,080,151	$690,167	$14,770,318

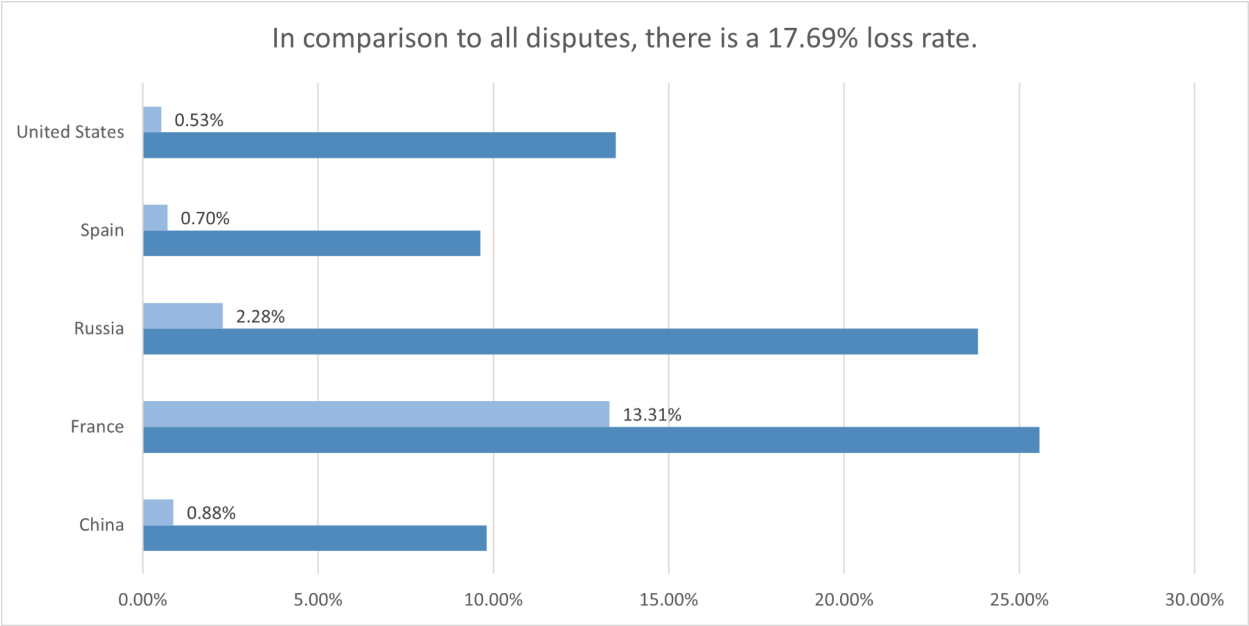
Findings:



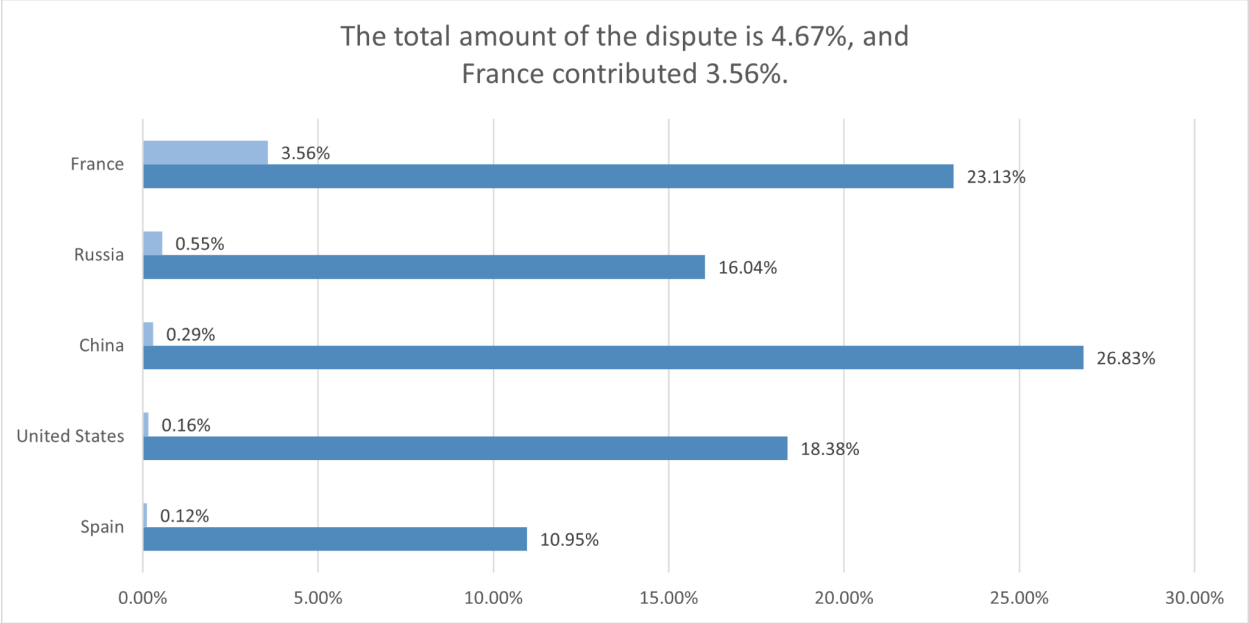
Russia has the longest average processing time of 29 days; while China has the shortest, with an average of 23 days. The average processing time with the whole dataset is 26 days.



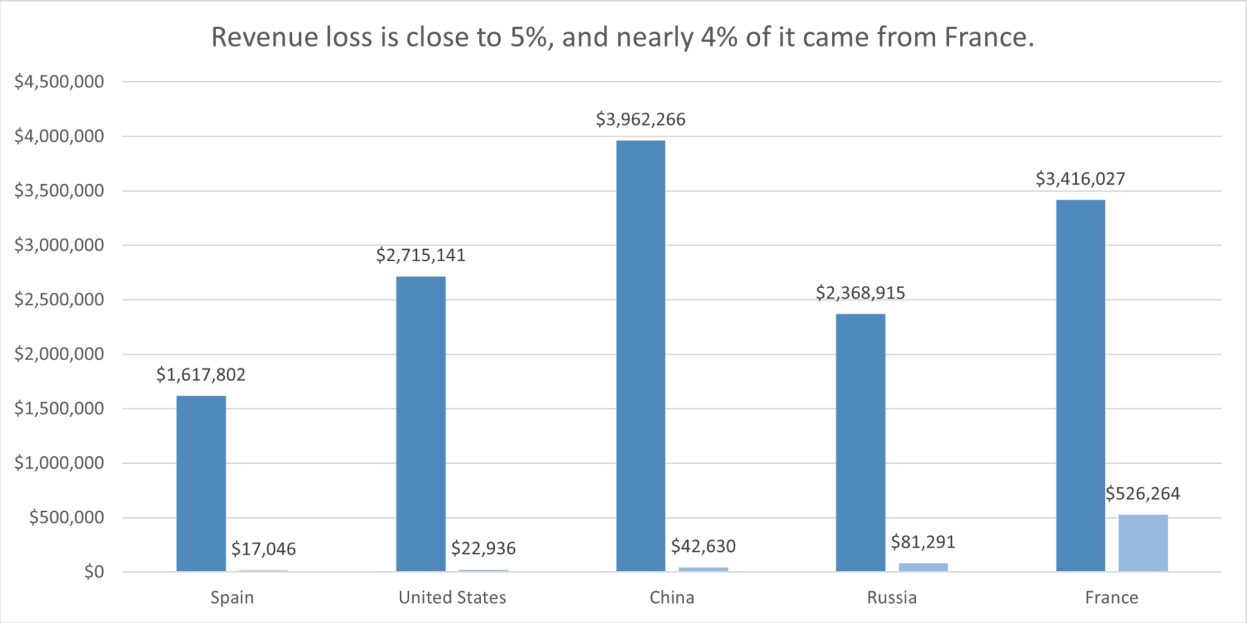
The United States has the longest average processing time to settle disputes, with an average of 41 days. France has the shortest average processing time to settle disputes, with an average of 34 days. The average processing time to settle disputes with the whole dataset is 36 days.



France has the highest percentage of lost disputes, with a value of 34.23%. While the United States has the lowest percentage of lost disputes at 3.75%.



France has the highest percentage of revenue lost, with a value of 13.35%. While the United States has the lowest percentage of revenue lost, with a value of 0.84%.



France has the highest invoice amount lost in disputes, with a value of $526,264. While Spain has the lowest invoice amount lost in disputes, with a value of $17,046.

Recommendations:

1. Hire legal experts
    - Hiring legal experts can provide guidance on legal matters, help draft legally sound contracts, keep the company compliant with regulations, and assist in navigating legal disputes. It's a valuable investment to prevent legal issues and provide guidance in the event of a dispute.
2. Review contract terms
    - Revisiting contracts with their clients, the company can investigate on what part of the agreement the client must have any misinterpretation or misunderstanding. Reviewing all the terms ensures that both parties involved are in agreement with what is expected of them, and protects both sides from any potential issues down the line.
3. Fast-track dispute resolution
    - Fast-tracking dispute resolution means solving disagreements quickly. This can save money, keep the company's good name, help us focus on important work, and make sure we follow the rules.
4. Modify payment terms
    - Avoiding revenue losses, the company can modify their payment terms and require down payments or deferred payments based on milestones achieved per project, in that way they can already see if they are losing or gaining profits.
5. Strategize compensation
    - Thinking of a new compensation strategy in order to lessen or mitigate the disputes from their clients to pave way in lessening the amounts lost.
6. Formulate case study

- Formulating a case study can provide valuable insights and understanding of the market by exploring various factors that influenced or contributed to the situation. The results of the case study can be used to draw conclusions and recommendations that can speed up the decision making process, resulting in a favorable resolution of the problem.

Additional Resources:

- Excel Workbook
- SQL Codes