

# STRIPE TASK

## DATA ANALYST INTERN

PRESENTED BY: PRIYANKA DEEP



-  Invoicing
-  Billing
-  Revenue Recognition
-  Revenue Reporting
-  Tax
-  Sigma
-  Data Pipeline



## ABOUT ME

As a performance-driven engineering professional, I focus on refining processes, maximizing efficiency, and enhancing user experiences through advanced development services. I specialize in process optimization, efficiency enhancements, and user experiences, collaborating with cutting-edge technologies for business transformation. My expertise includes data analysis, statistical modeling, and machine learning for managing data pipelines, conducting analyses, and solving complex problems to drive organizational success.

# Table of Contents

01

Problem Statement

02

Dataset Overview

03

Q1: Analysis, Rationale, Approach,  
Results

04

Q2: Analysis, Rationale, Approach,  
Results

05

Q3: Analysis

06

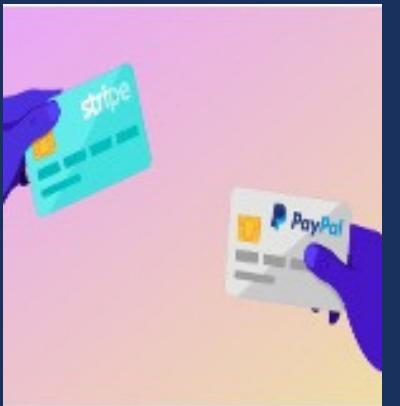
Conclusion

# PROBLEM STATEMENT

Stripe Connect is a product that allows platform businesses to pay out money to other businesses all over the world. In this project, you will analyze the payout behavior of Stripe platforms. A platform is a Stripe user that enables other businesses to process payments with Stripe.



# DATASET OVERVIEW



## Payout.csv

---

It has daily payout amounts (in cents) from platforms to recipients.



## Countries.csv

---

It maps merchants to countries. It has two columns, namely, merchant\_id and country.



## Industries.csv

---

It maps merchants to industries. It has two columns, namely, merchant\_id and country.



## Note

---

Both platforms and recipients are represented by a merchant id. In the dataset, the platform merchant is specified by `platform\_id` and the recipient is specified by `recipient\_id`.

# QUESTION 1:

**USING THE DATA PROVIDED, PLEASE PROVIDE AN ESTIMATE FOR THE AMOUNT OF MONEY WE SHOULD EXPECT TO BE PAID OUT TO EACH COUNTRY ON JAN. 1, 2019 (THE DAY AFTER THE LAST DAY IN THE DATASET).**

## ANALYSIS:

Conducted an exploratory analysis on all 3 files to manage the duplicated and/or null values in the dataset.

In order to study the date column, I transformed it to a more usable format, by extracting just the date part.

## RATIONALE:

To estimate the average payout on 1st Jan, 2019 it is important for us to look at the data from 2018. We will leverage predictive analytics for this task where we will predict future behaviour based on the historical data provided to us. Our ultimate goal is to predict the behaviour and trends of payouts made on Tuesdays on 2018. This is because 1st Jan,2019 falls on a Tuesday. So studying the behaviour of payouts on an average tuesday would give us better insights.

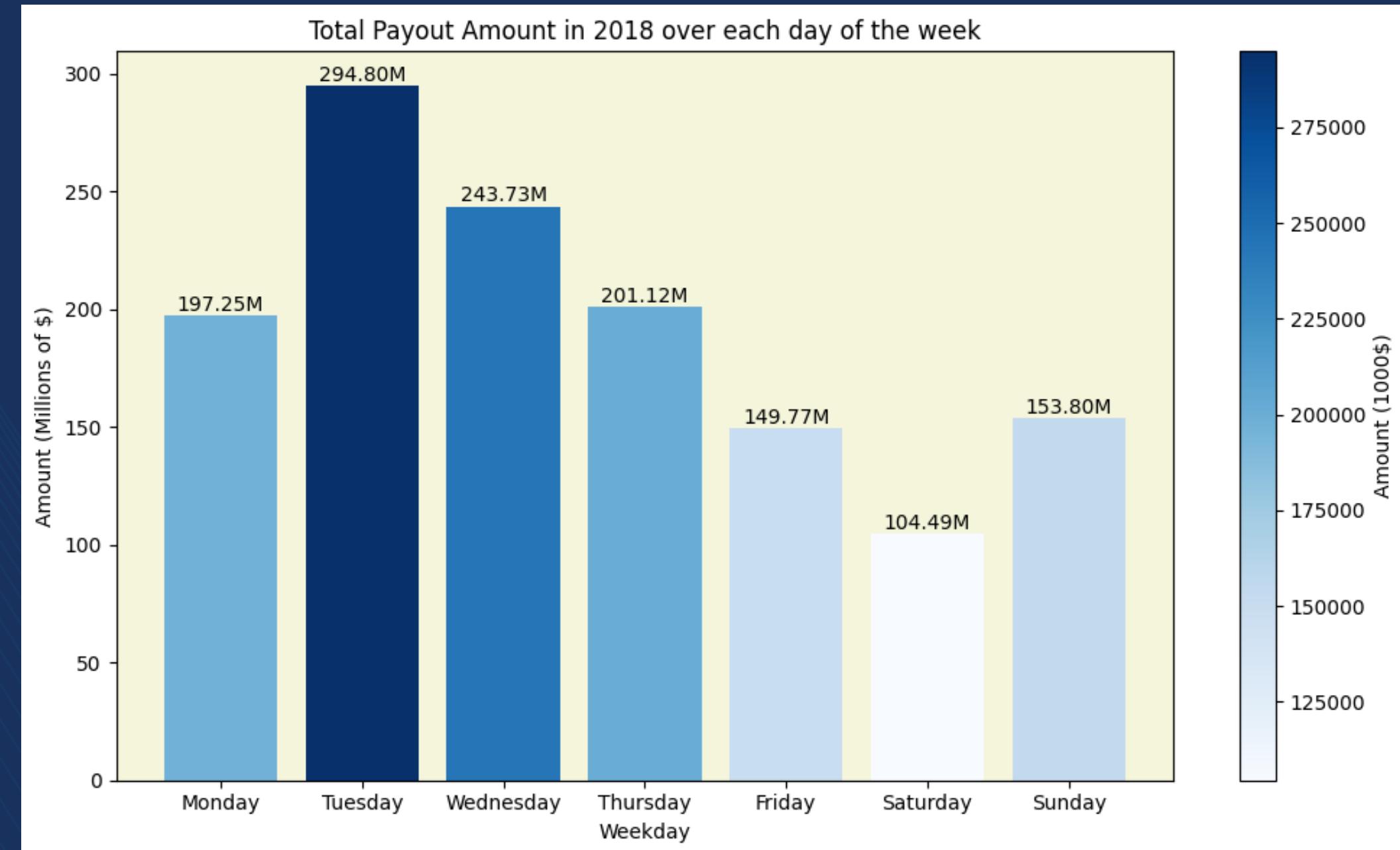
## APPROACH:

1. Identify the target day: 1st Jan is a Tuesday.
2. Data Transformation: Convert amount from cents to thousands. Change the date format for better understanding.
3. Identify the day of the payout using the date column.
4. Join and Merge: Using left join, merge the payouts and countries files while also removing redundant and unnecessary columns. Grouped the missing countries under new category "Unknown".
5. Data Filtering: On the new merged data, use a filter to keep only the payouts made on a Tuesday of 2018 and sum the amount to get an annual amount.
6. Now divide the summed amount by '52' to get the average payout value on a Tuesday(one year has 52 tuesdays)
7. Historical analysis on the generated data.

# PRELIMINARY INSIGHTS:

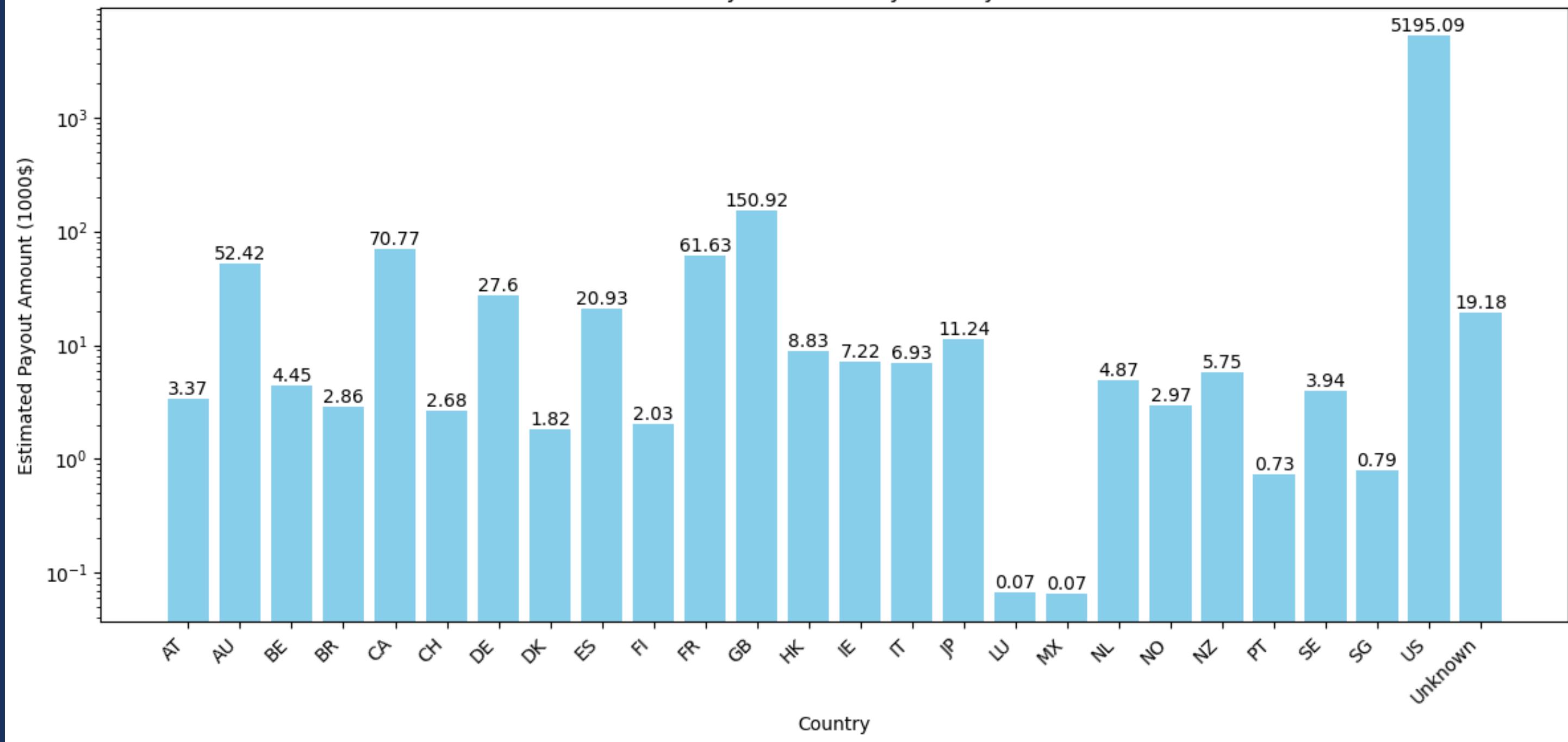
In the first graph, we can see a plot between the Total amount of payouts made in 2018 by the day of week. It is clear that Tuesdays accounted for the maximum payouts, approaching a total of about \$295 million. This gives us the following understanding:

- Tuesdays might be the peak transaction day for Stripe Connect. This could be due to billing cycles or consumer pattern.
- This implies dense business volume on Tuesdays which could be due to increased sales, transactions and financial activities during this day of the week.
- It also reflects a broader industry trend where Tuesday's are particularly active days for Stripe's clients.



# RESULTS :

Estimated Payout Amount by Country in 2019



The payouts that are estimated to be paid on 1st Jan 2019 according to the country indicates that the US has the highest payout amount of about \$5.2 million. The graph has been transformed to a logarithmic scale for better visualization.

This gives us an understanding of Stripe Connect's clientele which is majorly based in the US as it accounts for about 91% of the total payouts that are to be made to the other countries on 1st Jan, 2019.

This indicates the following:

1. Strong market dominance in the US. Least in Mexico.
2. Understanding the concentration of payouts in the US can help Stripe Connect evaluate its market penetration strategies. It may indicate areas for potential growth or diversification into other regions to reduce reliance on a single market.

Note :This estimate does not take seasonal variations into account so there is still scope for more accuracy.

# QUESTION 2:

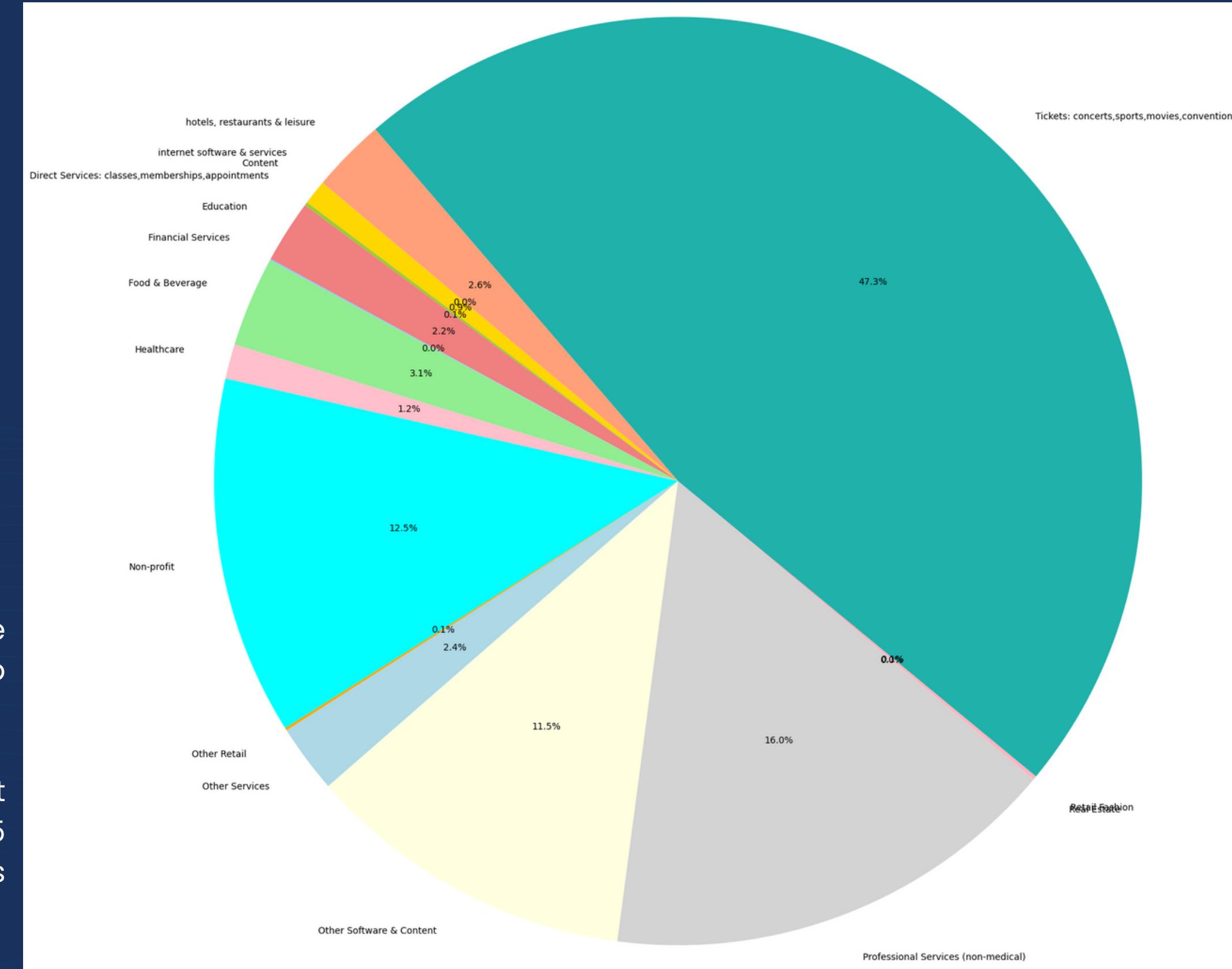
ASSUME THAT ONE YEAR FROM THE END OF THE DATASET, WE WILL HAVE 15 PLATFORMS FROM THE EDUCATION INDUSTRY, 5 FROM HOTELS, RESTAURANTS & LEISURE AND 40 FROM FOOD & BEVERAGE. HOW MUCH TOTAL PAYOUT VOLUME DO YOU EXPECT TO SEE ON A TYPICAL DAY IN 2019, FROM THESE THREE INDUSTRIES?

## RATIONALE:

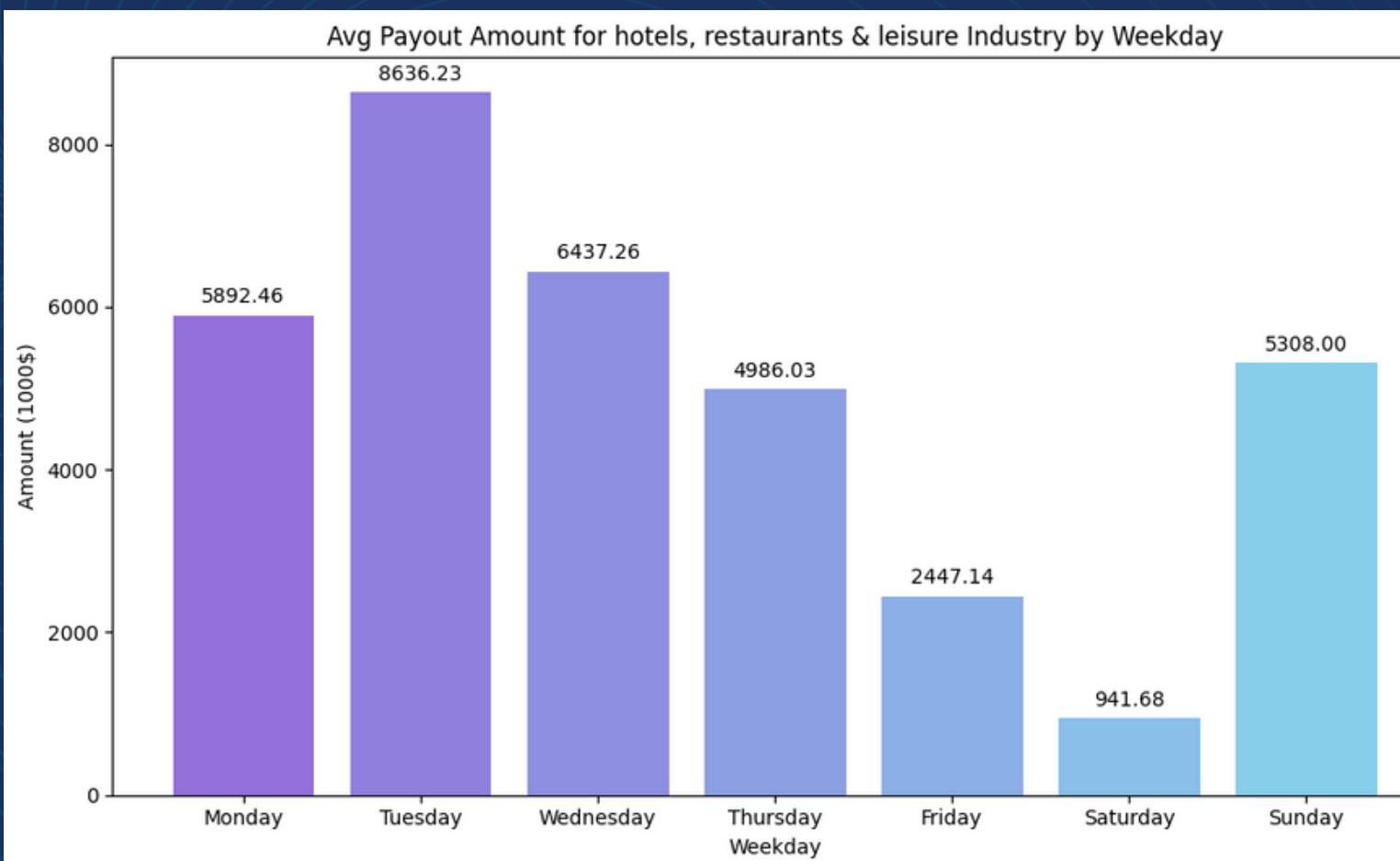
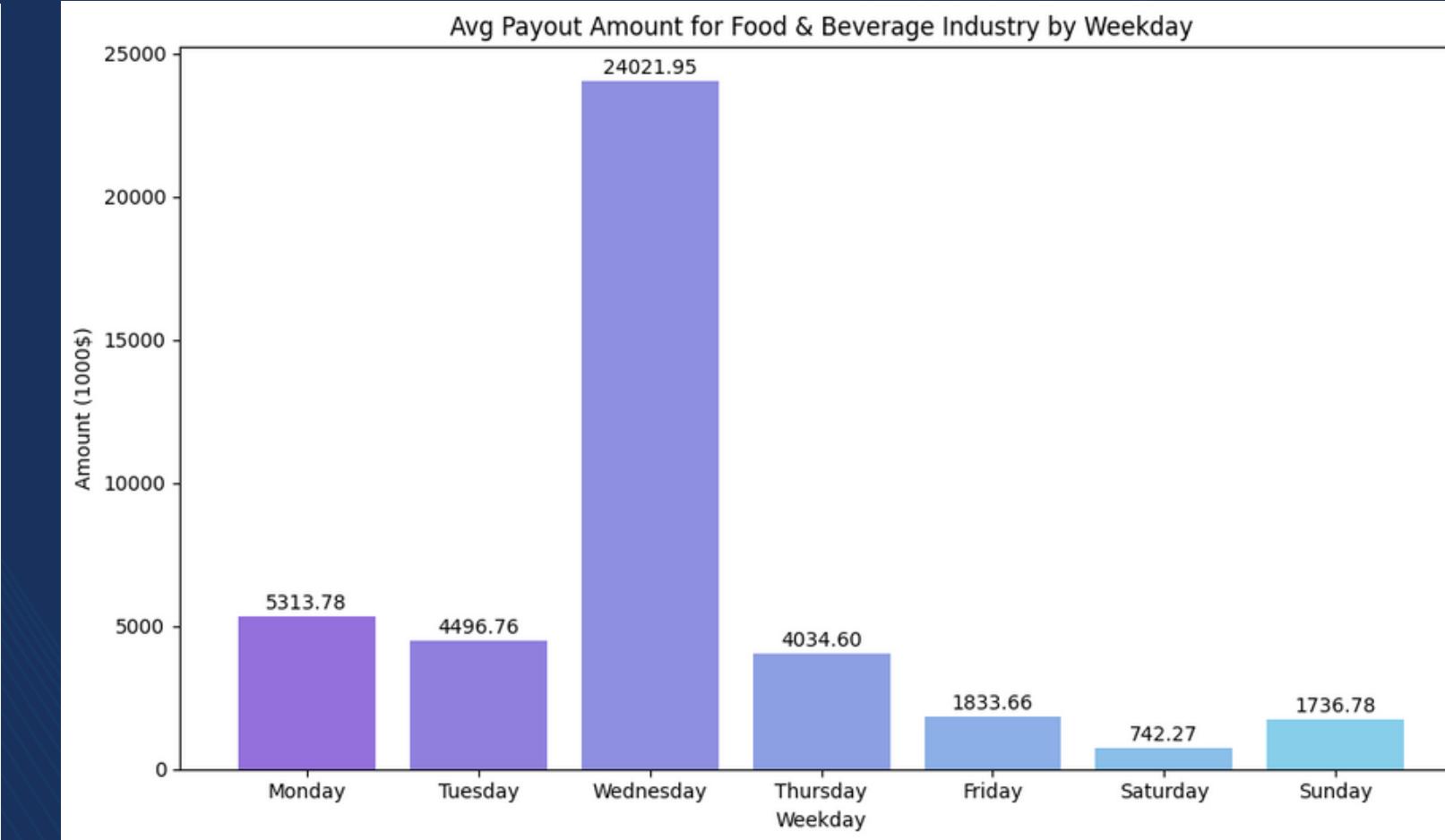
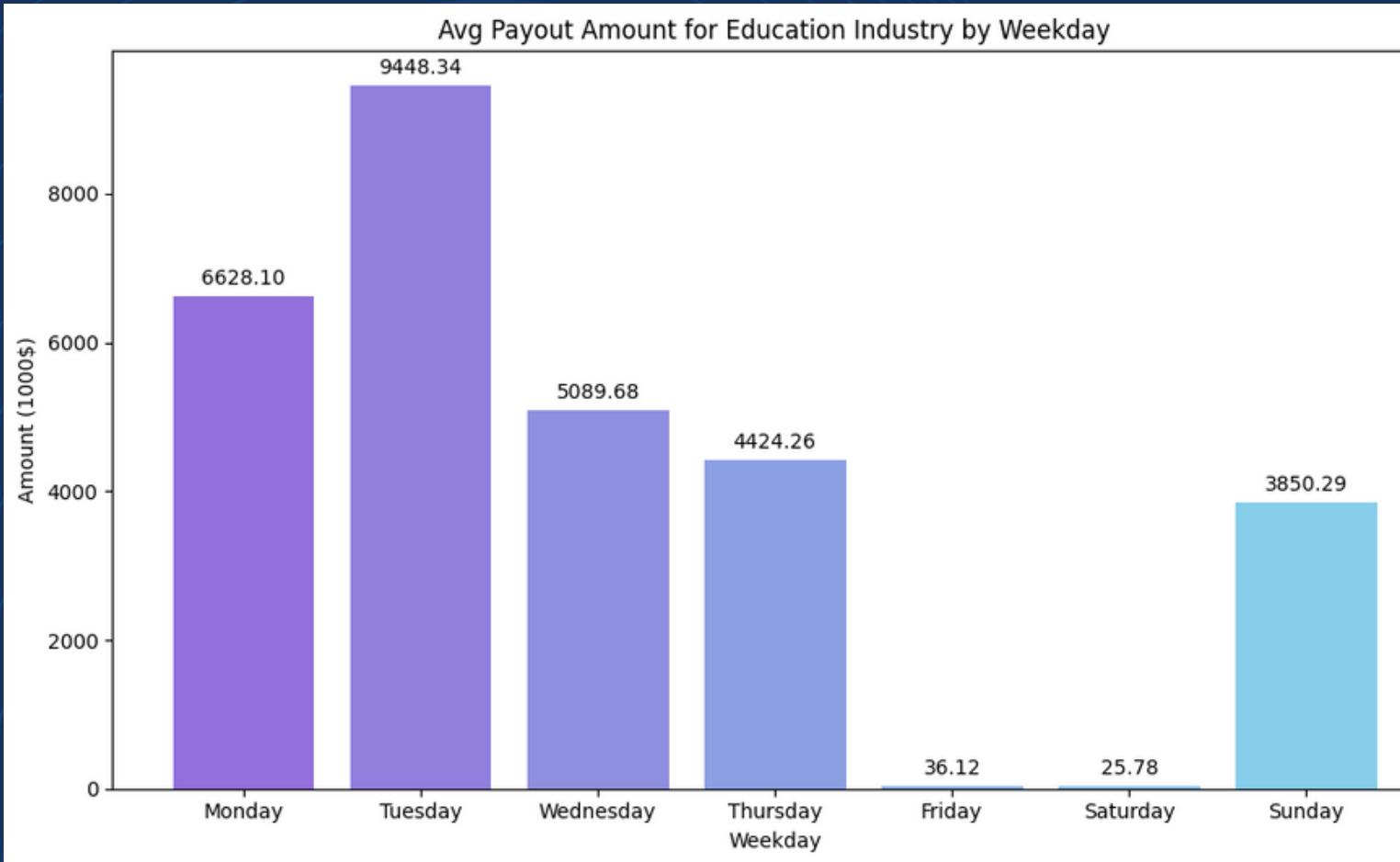
In order to estimate the total payout volume on a typical day in 2019, it is crucial for us to understand the payout in 2018 on each day. We will average the amount that is being paid each day in 2018 and based on that estimate the payout in 2019. However, we need to take into account that there is a change in the platform numbers for these 3 industries. So we also need to multiply the 2018 daily payout amount with a growth coefficient.

## APPROACH:

1. Regrouping outliers into relevant industry groups.
2. Using left join, merge the payouts and industries file. The missing values account for a very less sector so we drop them.
3. Calculate the payout amount on each day by the industry.
4. Estimate the expected daily payout value by multiplying it with the growth coefficient. For eg, education industry had 5 platforms already but it was expected to grow to 15. So its growth coefficient was 3(15/5).



# PRELIMINARY INSIGHTS:



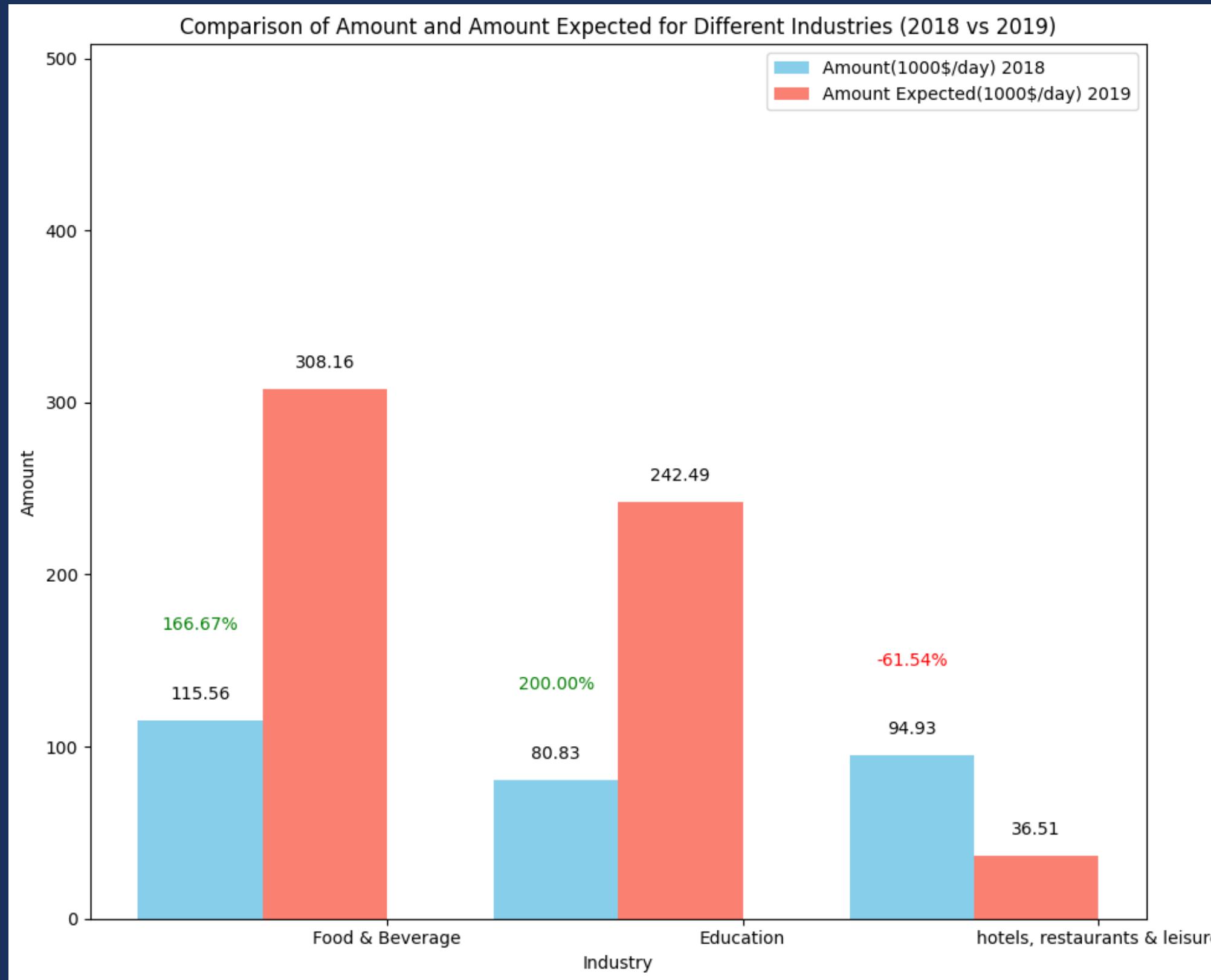
Given these graphs, we can see how the payout behaviour of these 3 different industries through the week.

The Education industry and the Hotels industry performed their best on Tuesdays while the Food & Beverage industry performed its best on Wednesdays.

This trend suggests that the industries experiences higher financial activity during weekdays, especially on Tuesday and Wednesdays, and significantly reduced activity during weekends.

The hotel, restaurants industry has a significant increase on Sundays which could indicate customer behaviour on a holiday.

# RESULTS:



Here is a graph that suggests a comparison of average daily payout volume in 2018 vs expected payout in 2019.

For the food & beverages industry, there is a sharp increase of about 166.67% in 2019, the education industry sees an increase of 200% in 2019 while the hotels industry will face a decrease of 61.54% in 2019.

The increase in the Food and Beverage and the Education industry indicate a strategic move for market expansion, perhaps. The daily payouts in 2018 were still high indicating positive performance even with fewer platforms. Increasing the platforms would be better justified with market analysis and revenue projections.

The decrease in the Hotel industry could indicate market trends and/or competition.

## QUESTION 3:

GIVEN THIS DATA, WHAT METRICS WOULD YOU DEFINE AND TRACK IN ORDER TO HELP PRODUCT AND BUSINESS TEAMS UNDERSTAND THE PROGRESS AND BEHAVIOR OF STRIPE CONNECT?

1. **Number of platform\_id** : The total number of connected accounts on the Stripe Connect's platform would help understand and track the adoption and growth of the platform.
2. **Transaction history**: To track the financial activity on the platform, it is important to look at metrics such as daily or weekly payouts. Just as in this project we were able to predict an estimate payout on a particular day of the year, we will be able to predict a lot more using this data. Demand forecasting, maintenance days, etc
3. **Transaction fees**: To accurately account for the finances, it is also important to know the fees that are generated upon usage.
4. **user interaction**: Metrics such as number of logins, time of login, clicks, API calls, would help us understand user behaviour and if necessary, make the interaction more engaging.
5. **Conversion rate**: It is also a crucial KPI as it enables us to understand a potential area of improvement. This metric helps in culminating better marketing campaigns or improve the UI/UX of the platform.
6. **Customer support metrics**: Metrics such as response time or customer satisfaction scores would help analyze the resolution rates better and also foster better retention rates.

# CONCLUSION:

These metrics and analysis of trends over time enables product and business teams to gain valuable insights into the performance, adoption, user behavior, and overall health of Stripe Connect. These insights can drive strategic decisions, product improvements, and growth strategies for the platform.



# THANK YOU

priyanka.deep98@gmail.com

[linkedin.com/in/priyanka-deep](https://www.linkedin.com/in/priyanka-deep)

206-841-5856