

Final Project - First Visualization

Programming for Business Analytics (11410ISS 406600)

Group 10

2025-11-03

Group Member

Choonhanunt Amnuaychaikij 111006224 Panuvit Tuicharoen 111006105 Napat Leesaksakul 112006208
Thanutchra Amnajcharoensak 112006412 Nantachaporn Sudjai 112006431

Question statement:

What is the most popular payment method among customers in the United Kingdom (UK) and the United States (USA) based on recent e-commerce transaction data?

Library Setup

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    4.0.0      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(stringr)
library(dplyr)
library(ggplot2)
library(lubridate)
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 4.5.2
```

Data Importing

```
ecommerce <- read.csv('ecommerce_dataset_10000.csv')
```

```
head(ecommerce)
```

```
##   customer_id first_name last_name gender age_group signup_date country
## 1   CUST2353     Erica   Oliver Female Teenagers  2022-06-29  Canada
## 2   CUST4463 Christopher   White   Male   Adults  2023-08-24   China
## 3   CUST4512   Spencer   Foster   Male   Senior  2023-07-18 Germany
## 4   CUST5711   Jessica   Harris   Male Teenagers  2025-08-22  France
## 5   CUST1296     Amy    Johnson Female Teenagers  2021-03-23  Brazil
## 6   CUST2790   Shelby   Sutton  Other   Adults  2025-07-18  Canada
##   product_id   product_name      category quantity unit_price order_id
## 1   PROD108     Fitbit Versa 3   Electronics      3         229 ORD10000
## 2   PROD103     Levi's Jeans      Apparel      4          59 ORD10001
## 3   PROD111   Lego Star Wars Set      Toys      2          59 ORD10002
## 4   PROD107     Dyson Vacuum Home & Kitchen      4         399 ORD10003
## 5   PROD105 Adidas Running Shoes      Apparel      1         110 ORD10004
## 6   PROD108     Fitbit Versa 3   Electronics      5         229 ORD10005
##   order_date order_status payment_method rating review_text review_id
## 1 2023-07-13     Pending      Credit Card      2      good  REV20000
## 2 2024-08-12     Pending      PayPal          2    average  REV20001
## 3 2024-08-04   Delivered Cash on Delivery      5      good  REV20002
## 4 2025-05-23   Delivered Cash on Delivery      2    very good  REV20003
## 5 2023-07-02     Returned Cash on Delivery      1    very good  REV20004
## 6 2023-04-13     Returned      PayPal          3    very good  REV20005
##   review_date
## 1 2025-06-06
## 2 2023-08-05
## 3 2023-01-03
## 4 2023-03-14
## 5 2023-10-18
## 6 2023-02-14
```

Data Cleaning

```
ecommerce <- ecommerce %>%
  mutate(
    order_year = year(ymd(order_date))
  )
summary(ecommerce$order_year)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2022    2023    2024    2024    2024    2025
```

```
ecommerce_clean <- ecommerce %>%
  filter(
    !is.na(payment_method),
```

```

country %in% c("UK", "USA"),
order_year %in% c(2022, 2024)
)

head(ecommerce_clean)

```

```

##   customer_id first_name last_name gender age_group signup_date country
## 1   CUST2451   Barbara   Hansen Female   Adults   2024-11-10      UK
## 2   CUST1438  Michelle   Vargas   Male   Adults   2023-07-11      UK
## 3   CUST2997   Amanda   Martinez Female   Senior   2021-06-02      USA
## 4   CUST2895  Lawrence   Hines   Female   Senior   2021-03-15      USA
## 5   CUST1182    John    Jacobs   Male   Senior   2023-03-09      UK
## 6   CUST4751   Tyler    Martin   Male   Adults   2020-11-07      USA
##   product_id   product_name   category quantity unit_price order_id
## 1   PROD103     Levi's Jeans     Apparel      2         59 ORD10007
## 2   PROD109   Kindle Paperwhite   Books      1        129 ORD10012
## 3   PROD105 Adidas Running Shoes   Apparel      3        110 ORD10025
## 4   PROD102   Sony Headphones Electronics    5        199 ORD10029
## 5   PROD113 Wilson Tennis Racket   Sports      3        149 ORD10035
## 6   PROD112   Barbie Dreamhouse    Toys      3        199 ORD10047
##   order_date order_status   payment_method rating review_text review_id
## 1 2024-01-29   Pending     Credit Card      1   very good   REV20007
## 2 2024-05-15   Shipped     Credit Card      1    very bad   REV20012
## 3 2024-05-21   Pending Cash on Delivery      2    average   REV20025
## 4 2024-12-31   Cancelled   Credit Card      5   very good   REV20029
## 5 2022-12-11   Cancelled      PayPal      5    average   REV20035
## 6 2024-12-28   Cancelled   Credit Card      1    average   REV20047
##   review_date order_year
## 1 2025-06-02      2024
## 2 2024-12-28      2024
## 3 2023-06-05      2024
## 4 2024-12-04      2024
## 5 2024-02-20      2022
## 6 2024-12-07      2024

```

Calculate Most Popular Payment Method by Country

UK Most Popular Payment Method

```

payment_summary_uk <- ecommerce_clean %>%
  filter(country == "UK", order_year %in% c(2022, 2024)) %>%
  group_by(order_year, payment_method) %>%
  summarise(total_transactions = n(), .groups = "drop")

payment_summary_uk

```

```

## # A tibble: 6 x 3
##   order_year payment_method   total_transactions
##       <dbl> <chr>               <int>
## 1      2022 Cash on Delivery           30

```

```
## 2      2022 Credit Card      39
## 3      2022 PayPal          38
## 4      2024 Cash on Delivery 101
## 5      2024 Credit Card     103
## 6      2024 PayPal          102
```

```
top_methods_uk <- payment_summary_uk %>%
  group_by(payment_method) %>%
  summarise(overall = sum(total_transactions)) %>%
  slice_max(overall, n = 3) %>%
  pull(payment_method)
```

```
top_methods_uk
```

```
## [1] "Credit Card"      "PayPal"           "Cash on Delivery"
```

USA Most Popular Payment Method

```
payment_summary_usa <- ecommerce_clean %>%
  filter(country == "USA", order_year %in% c(2022, 2024)) %>%
  group_by(order_year, payment_method) %>%
  summarise(total_transactions = n(), .groups = "drop")
```

```
payment_summary_usa
```

```
## # A tibble: 6 x 3
##   order_year payment_method total_transactions
##   <dbl> <chr>                <int>
## 1      2022 Cash on Delivery      42
## 2      2022 Credit Card          49
## 3      2022 PayPal              33
## 4      2024 Cash on Delivery     126
## 5      2024 Credit Card         117
## 6      2024 PayPal              98
```

```
top_methods_usa <- payment_summary_usa %>%
  group_by(payment_method) %>%
  summarise(overall = sum(total_transactions)) %>%
  slice_max(overall, n = 3) %>%
  pull(payment_method)
```

```
top_methods_usa
```

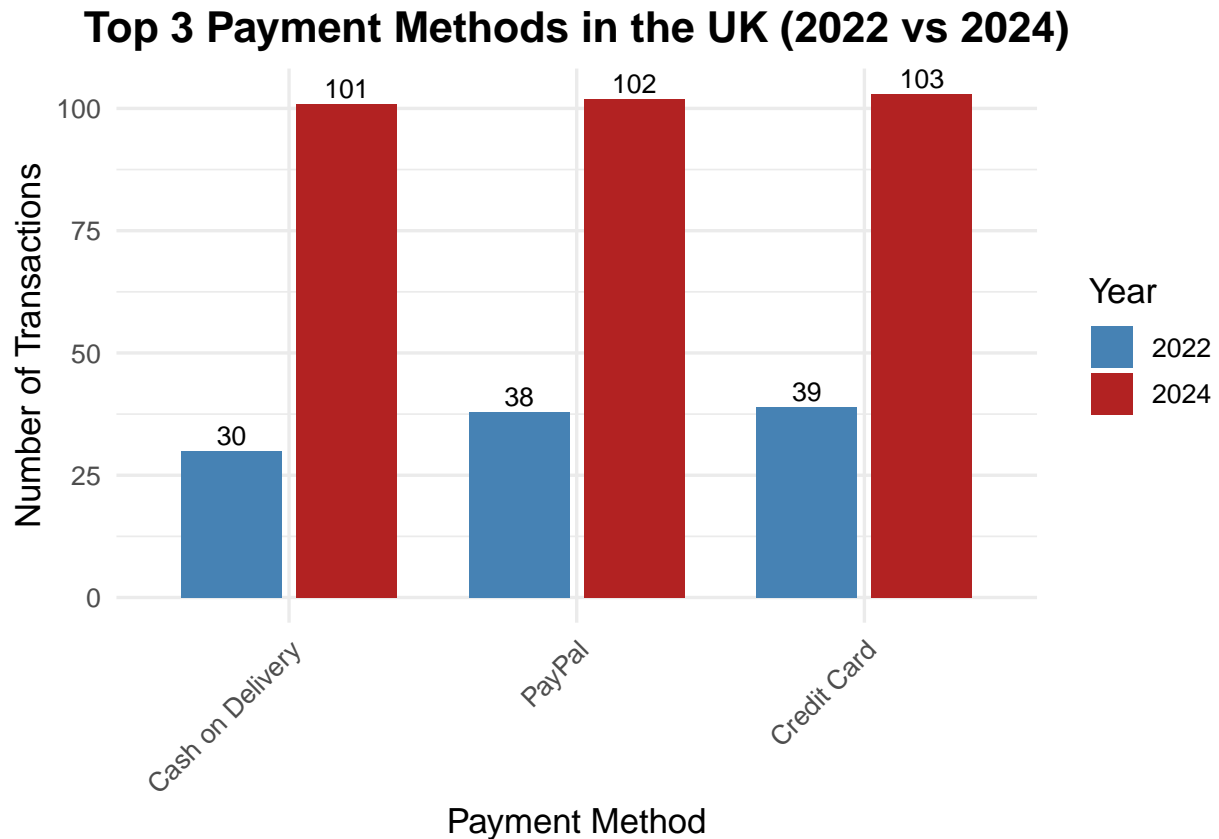
```
## [1] "Cash on Delivery" "Credit Card"     "PayPal"
```

Visualization Section

UK Visualization

```
payment_summary_uk <- payment_summary_uk %>%
  filter(payment_method %in% top_methods_uk)

ggplot(payment_summary_uk,
  aes(x = reorder(payment_method, total_transactions, FUN = sum),
    y = total_transactions,
    fill = as.factor(order_year),
    label = total_transactions)) +
  geom_col(position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(position = position_dodge(width = 0.8),
    vjust = -0.3, size = 3.5) +
  scale_fill_manual(
    values = c("2022" = "steelblue", "2024" = "firebrick"),
    name = "Year"
  ) +
  labs(
    title = "Top 3 Payment Methods in the UK (2022 vs 2024)",
    x = "Payment Method",
    y = "Number of Transactions"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
)
```



The bar chart illustrates the top three payment methods used by UK customers in 2022 and 2024. Across all methods, transaction volumes increased substantially between the two years, approximately a threefold growth, suggesting expanding e-commerce activity and customer adoption.

Among the payment options, Credit Card emerged as the most frequently used method, maintaining a clear lead in both years. PayPal ranked second, followed by Cash on Delivery as the least common option. This trend suggests that UK consumers continue to prefer digital and card-based payments, reflecting high trust in secure, cashless transactions and widespread access to credit systems.

USA Visualization

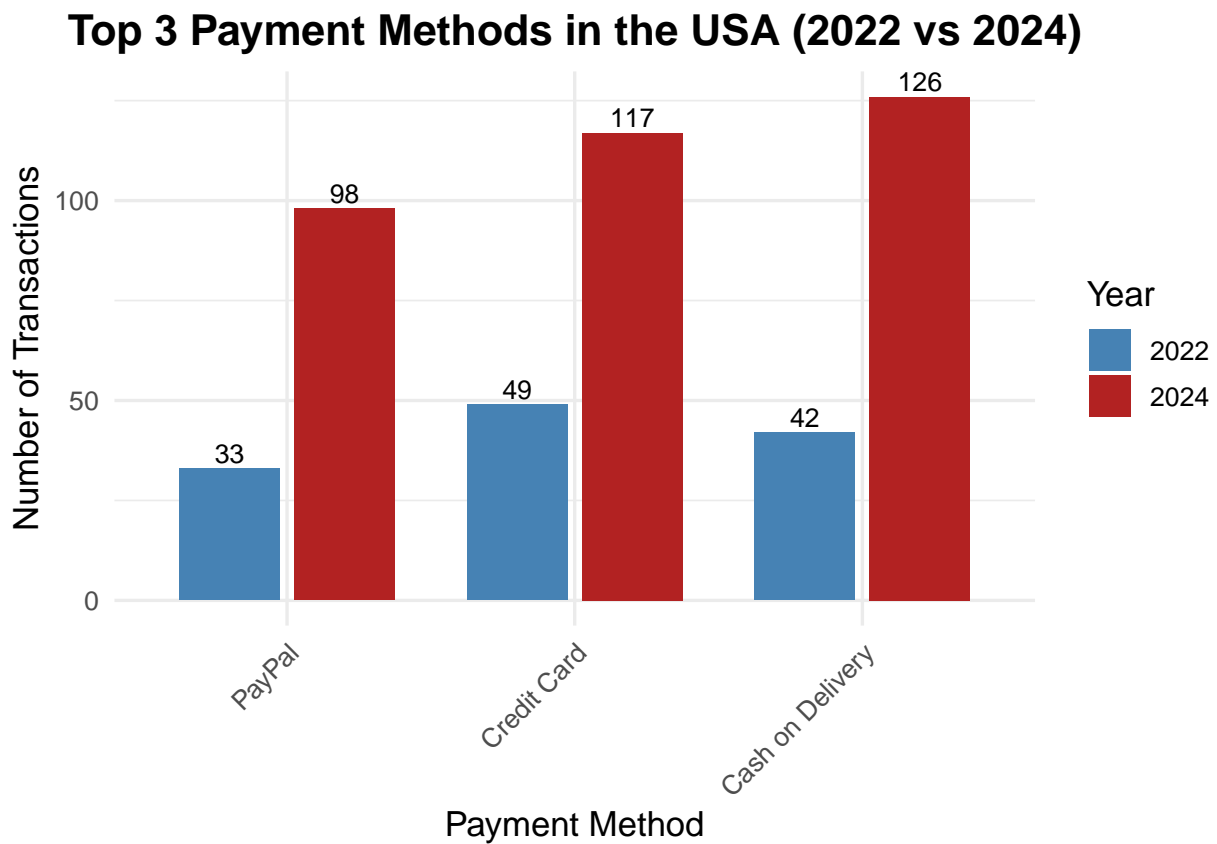
```
payment_summary_usa <- payment_summary_usa %>%
  filter(payment_method %in% top_methods_usa)

ggplot(payment_summary_usa,
  aes(x = reorder(payment_method, total_transactions, FUN = sum),
    y = total_transactions,
    fill = as.factor(order_year),
    label = total_transactions)) +
  geom_col(position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(position = position_dodge(width = 0.8),
    vjust = -0.3, size = 3.5) +
  scale_fill_manual(
    values = c("2022" = "steelblue", "2024" = "firebrick"),
    name = "Year"
```

```

) +
labs(
  title = "Top 3 Payment Methods in the USA (2022 vs 2024)",
  x = "Payment Method",
  y = "Number of Transactions"
) +
theme_minimal(base_size = 13) +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold"),
  axis.text.x = element_text(angle = 45, hjust = 1)
)

```



The chart displays the top three payment methods among U.S. customers in 2022 and 2024. While Credit Card was the dominant method in 2022, by 2024 Cash on Delivery had overtaken it, showing a notable shift in consumer behavior.

Meanwhile, PayPal consistently remained the least used method, though it also experienced growth between the two years. Overall, U.S. consumers appear to balance convenience and control with card payments remaining strong but Cash on Delivery gaining traction as a trusted alternative.