Airline Pilot Retention Analysis

Mike Hickey

2025-05-14

Table of Contents

raw <- read\_csv(  
 here(params$data\_file), # data/retention.csv  
 show\_col\_types = FALSE  
)  
  
# Row 1 = headers; rows 2‑3 are survey text & metadata → drop them  
retention\_data <- raw %>%  
 slice(-(1:2)) %>% # remove rows 2 & 3  
 clean\_names() # snake\_case headers  
  
# Remove unwanted metadata columns  
ignore\_cols <- c(  
 "start\_date", "end\_date", "status", "progress", "duration\_in\_seconds",  
 "recorded\_date", "response\_id", "distribution\_channel",  
 "user\_language", "q\_recaptcha\_score"  
)  
retention\_data <- retention\_data %>% select(-all\_of(ignore\_cols))  
  
glimpse(retention\_data)

## Rows: 113  
## Columns: 52  
## $ finished <chr> "False", "False", "False", "False", "True", "True"…  
## $ informed\_consent <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "…  
## $ screener <chr> "Yes", "Yes", NA, NA, "Yes", "Yes", "Yes", "Yes", …  
## $ age <chr> NA, NA, NA, NA, "27", "28", "48", "34", "23", "42"…  
## $ gender <chr> NA, NA, NA, NA, "Male", "Male", "Female", "Male", …  
## $ nationality <chr> NA, NA, NA, NA, "United States of America", "Unite…  
## $ base\_state <chr> NA, NA, NA, NA, "California", "Ohio", "Montana", "…  
## $ position\_position <chr> NA, NA, NA, NA, "First Officer", "Captain", "Capta…  
## $ air\_carrier\_type <chr> NA, NA, NA, NA, "Low Cost Carrier", "Regional Airl…  
## $ experience <chr> NA, NA, NA, NA, "3", "3", "11", "4", "1", "2", "8"…  
## $ total\_flying\_hours <chr> NA, NA, NA, NA, "3000", "2800", "6800", "3600", "2…  
## $ military\_experience <chr> NA, NA, NA, NA, "Never served in the military", "N…  
## $ general\_1 <chr> NA, NA, NA, NA, "2", "1", "1", "1", "1", "1", "2",…  
## $ general\_2 <chr> NA, NA, NA, NA, "1", "4", "2", "2", "2", "6", "4",…  
## $ general\_3 <chr> NA, NA, NA, NA, "6", "2", "5", "3", "4", "3", "6",…  
## $ general\_4 <chr> NA, NA, NA, NA, "5", "5", "6", "6", "6", "2", "5",…  
## $ general\_5 <chr> NA, NA, NA, NA, "3", "3", "3", "4", "3", "4", "1",…  
## $ general\_6 <chr> NA, NA, NA, NA, "4", "6", "4", "5", "5", "5", "3",…  
## $ financial\_1 <chr> NA, NA, NA, NA, "1", "2", "1", "1", "1", "1", "1",…  
## $ financial\_2 <chr> NA, NA, NA, NA, "2", "5", "2", "4", "4", "3", "5",…  
## $ financial\_3 <chr> NA, NA, NA, NA, "4", "4", "3", "2", "2", "4", "2",…  
## $ financial\_4 <chr> NA, NA, NA, NA, "3", "3", "5", "6", "6", "5", "4",…  
## $ financial\_5 <chr> NA, NA, NA, NA, "5", "1", "4", "3", "3", "2", "3",…  
## $ financial\_6 <chr> NA, NA, NA, NA, "6", "6", "6", "5", "5", "6", "6",…  
## $ qo\_l\_1 <chr> NA, NA, NA, NA, "1", "3", "1", "3", "2", "1", "2",…  
## $ qo\_l\_2 <chr> NA, NA, NA, NA, "2", "6", "3", "4", "5", "5", "4",…  
## $ qo\_l\_3 <chr> NA, NA, NA, NA, "4", "4", "6", "6", "6", "6", "6",…  
## $ qo\_l\_4 <chr> NA, NA, NA, NA, "6", "2", "4", "1", "3", "2", "1",…  
## $ qo\_l\_5 <chr> NA, NA, NA, NA, "5", "5", "5", "5", "4", "3", "3",…  
## $ qo\_l\_6 <chr> NA, NA, NA, NA, "3", "1", "2", "2", "1", "4", "5",…  
## $ professional\_1 <chr> NA, NA, NA, NA, "1", "1", "1", "1", "2", "1", "1",…  
## $ professional\_2 <chr> NA, NA, NA, NA, "3", "4", "4", "2", "4", "2", "4",…  
## $ professional\_3 <chr> NA, NA, NA, NA, "4", "2", "3", "4", "3", "3", "5",…  
## $ professional\_4 <chr> NA, NA, NA, NA, "5", "5", "5", "3", "1", "5", "2",…  
## $ professional\_5 <chr> NA, NA, NA, NA, "2", "3", "2", "5", "5", "4", "3",…  
## $ recognition\_1 <chr> NA, NA, NA, NA, "2", "3", "3", "3", "2", "1", "4",…  
## $ recognition\_3 <chr> NA, NA, NA, NA, "1", "2", "2", "1", "3", "2", "1",…  
## $ recognition\_4 <chr> NA, NA, NA, NA, "3", "1", "1", "2", "1", "3", "3",…  
## $ recognition\_7 <chr> NA, NA, NA, NA, "4", "4", "4", "4", "4", "4", "2",…  
## $ schedule\_1 <chr> NA, NA, NA, NA, "5", "1", "1", "2", "5", "1", "1",…  
## $ schedule\_2 <chr> NA, NA, NA, NA, "3", "5", "4", "3", "3", "4", "5",…  
## $ schedule\_3 <chr> NA, NA, NA, NA, "1", "2", "5", "1", "2", "2", "2",…  
## $ schedule\_4 <chr> NA, NA, NA, NA, "4", "4", "2", "4", "1", "3", "4",…  
## $ schedule\_5 <chr> NA, NA, NA, NA, "2", "3", "3", "5", "4", "5", "3",…  
## $ operational\_1 <chr> NA, NA, NA, NA, "1", "1", "1", "3", "1", "3", "3",…  
## $ operational\_2 <chr> NA, NA, NA, NA, "2", "2", "4", "5", "2", "1", "1",…  
## $ operational\_3 <chr> NA, NA, NA, NA, "3", "4", "2", "1", "3", "2", "2",…  
## $ operational\_4 <chr> NA, NA, NA, NA, "5", "5", "5", "2", "5", "5", "5",…  
## $ operational\_5 <chr> NA, NA, NA, NA, "4", "3", "3", "4", "4", "4", "4",…  
## $ leaving <chr> NA, NA, NA, NA, "Maybe", "Yes", "Yes", "Yes", "Yes…  
## $ when <chr> NA, NA, NA, NA, "1-2 years", "1-2 years", "Within …  
## $ why <chr> NA, NA, NA, NA, "Bases in lower cost of living are…

total\_records <- nrow(retention\_data)  
  
cleaned <- retention\_data %>%  
 mutate(  
 finished = toupper(as.character(finished)),  
 informed\_consent = toupper(as.character(informed\_consent)),  
 screener = toupper(as.character(screener))  
 ) %>%  
 mutate(  
 exclusion\_reason = case\_when(  
 informed\_consent != "YES" ~ "did\_not\_consent",  
 finished != "TRUE" ~ "did\_not\_finish",  
 screener != "YES" ~ "not\_air\_carrier",  
 TRUE ~ "included"  
 )  
 )  
  
exclusion\_table <- cleaned %>%  
 count(exclusion\_reason, name = "n")  
  
knitr::kable(exclusion\_table, caption = "Why Records Were Excluded")

Why Records Were Excluded

| exclusion\_reason | n |
| --- | --- |
| did\_not\_finish | 35 |
| included | 76 |
| not\_air\_carrier | 2 |

included\_data <- cleaned %>%  
 filter(exclusion\_reason == "included") %>%  
 select(-exclusion\_reason) # drop helper column  
  
cat("\*\*Total records in file:\*\*", total\_records,  
 "\\n\\n\*\*Records kept for analysis:\*\*", nrow(included\_data),  
 "\\n\\n\*\*Records excluded:\*\*", total\_records - nrow(included\_data))

## \*\*Total records in file:\*\* 113 \n\n\*\*Records kept for analysis:\*\* 76 \n\n\*\*Records excluded:\*\* 37

write\_csv(included\_data, here("data/retention\_cleaned.csv"))

Preview of Cleaned Data

| finished | informed\_consent | screener | age | gender | nationality | base\_state | position\_position | air\_carrier\_type | experience | total\_flying\_hours | military\_experience | general\_1 | general\_2 | general\_3 | general\_4 | general\_5 | general\_6 | financial\_1 | financial\_2 | financial\_3 | financial\_4 | financial\_5 | financial\_6 | qo\_l\_1 | qo\_l\_2 | qo\_l\_3 | qo\_l\_4 | qo\_l\_5 | qo\_l\_6 | professional\_1 | professional\_2 | professional\_3 | professional\_4 | professional\_5 | recognition\_1 | recognition\_3 | recognition\_4 | recognition\_7 | schedule\_1 | schedule\_2 | schedule\_3 | schedule\_4 | schedule\_5 | operational\_1 | operational\_2 | operational\_3 | operational\_4 | operational\_5 | leaving | when | why |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRUE | YES | YES | 27 | Male | United States of America | California | First Officer | Low Cost Carrier | 3 | 3000 | Never served in the military | 2 | 1 | 6 | 5 | 3 | 4 | 1 | 2 | 4 | 3 | 5 | 6 | 1 | 2 | 4 | 6 | 5 | 3 | 1 | 3 | 4 | 5 | 2 | 2 | 1 | 3 | 4 | 5 | 3 | 1 | 4 | 2 | 1 | 2 | 3 | 5 | 4 | Maybe | 1-2 years | Bases in lower cost of living areas. |
| TRUE | YES | YES | 28 | Male | United States of America | Ohio | Captain | Regional Airline | 3 | 2800 | Never served in the military | 1 | 4 | 2 | 5 | 3 | 6 | 2 | 5 | 4 | 3 | 1 | 6 | 3 | 6 | 4 | 2 | 5 | 1 | 1 | 4 | 2 | 5 | 3 | 3 | 2 | 1 | 4 | 1 | 5 | 2 | 4 | 3 | 1 | 2 | 4 | 5 | 3 | Yes | 1-2 years | For a better airline that has better opportunities |
| TRUE | YES | YES | 48 | Female | United States of America | Montana | Captain | Freight/Cargo Carrier | 11 | 6800 | Never served in the military | 1 | 2 | 5 | 6 | 3 | 4 | 1 | 2 | 3 | 5 | 4 | 6 | 1 | 3 | 6 | 4 | 5 | 2 | 1 | 4 | 3 | 5 | 2 | 3 | 2 | 1 | 4 | 1 | 4 | 5 | 2 | 3 | 1 | 4 | 2 | 5 | 3 | Yes | Within 1 year | Unfair, unsafe, and fatiguing work rules/schedules. Company is not negotiating in good faith. |
| TRUE | YES | YES | 34 | Male | United States of America | Pennsylvania | Captain | Regional Airline | 4 | 3600 | Previously served in the military | 1 | 2 | 3 | 6 | 4 | 5 | 1 | 4 | 2 | 6 | 3 | 5 | 3 | 4 | 6 | 1 | 5 | 2 | 1 | 2 | 4 | 3 | 5 | 3 | 1 | 2 | 4 | 2 | 3 | 1 | 4 | 5 | 3 | 5 | 1 | 2 | 4 | Yes | 1-2 years | More pay, better 401k, better reserve schedule, higher seniority for non-reving, better ability to manipulate my scheudle. |
| TRUE | YES | YES | 23 | Male | United States of America | Ohio | First Officer | Regional Airline | 1 | 2400 | Never served in the military | 1 | 2 | 4 | 6 | 3 | 5 | 1 | 4 | 2 | 6 | 3 | 5 | 2 | 5 | 6 | 3 | 4 | 1 | 2 | 4 | 3 | 1 | 5 | 2 | 3 | 1 | 4 | 5 | 3 | 2 | 1 | 4 | 1 | 2 | 3 | 5 | 4 | Yes | 1-2 years | Transition to mainline, for better pay, and benefits |
| TRUE | YES | YES | 42 | Male | United States of America | Illinois | First Officer | Regional Airline | 2 | 2500 | Never served in the military | 1 | 6 | 3 | 2 | 4 | 5 | 1 | 3 | 4 | 5 | 2 | 6 | 1 | 5 | 6 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 4 | 1 | 2 | 3 | 4 | 1 | 4 | 2 | 3 | 5 | 3 | 1 | 2 | 5 | 4 | Maybe | 1-2 years | Fly larger aircraft that pays better |

## NEXT STEPS

Descriptive summaries (age, gender, etc.) Mann–Whitney U and Kruskal–Wallis tests Visualisations (box‑plots, violin plots, etc.)