Nice, this sample is **perfect** 👍 — it exposes *all the messy real-world OTT cases*: renewals, versioning, churn, winback, country change, overlaps.

I'll do this **step-by-step**, **using ONLY your provided data**, and I'll explain **what each retention analysis actually means in business terms**.

0️⃣ Load & Prepare the Data (Exactly as Given)

```
library(dplyr)
library(lubridate)
library(tidyr)

subs <- tibble::tribble(
  ~CustomerID, ~Country, ~Direct_Indirect, ~Product_name, ~Offer_period,
  ~Subscription_ID, ~Subscription_version, ~Subscription_latest,
  ~Subscription_status, ~Subscription_type,
  ~Subscription_start_date, ~Expiry_date,


101,"USA","Direct","XYZTotal","monthly",8889,1,0,"churned","churned","1/18/2025","2/18/20
25",

101,"USA","Direct","FootballSeasonal","monthly",8889,2,1,"active","new","1/18/2025","2/18/2
025",

101,"USA","Direct","XYZTotal","annual",2024,1,1,"active","winback","2/18/2025","2/18/2026",

102,"India","Direct","FootballSeasonal","annual",6752,1,1,"active","new","2/18/2025","2/18/2
026",

102,"Qatar","Direct","4KTotal","annual",1220,1,1,"active","winback","3/18/2025","4/18/2026",
  103,"USA","Direct","4KTotal","monthly",2390,1,0,"active","new","2/28/2025","3/28/2025",

103,"USA","Direct","4KTotal","monthly",2390,2,1,"churned","churned","2/28/2025","3/28/2025
",

103,"USA","Direct","4KTotal","monthly",2450,1,1,"active","winback","3/18/2025","4/18/2026",

103,"USA","Direct","4KTotal","monthly",2670,1,1,"active","winback","2/29/2025","3/29/2025",
  104,"Kenya","Direct","XYZTotal","annual",3498,1,0,"active","new","2/18/2025","3/18/2025",

104,"Kenya","Direct","XYZTotal","annual",3498,2,1,"churned","churned","2/18/2025","3/18/20
25"
) %>%
  mutate(
```

```
  Subscription_start_date = mdy(Subscription_start_date),
  Expiry_date = mdy(Expiry_date),
  start_month = floor_date(Subscription_start_date, "month")
 )
```

1️⃣ Current Retained Customers (Top-line KPI)

Business question:
👉 *How many customers are currently active?*

```
subs %>%
 filter(Subscription_latest == 1) %>%
 summarise(
  total_customers = n_distinct(CustomerID),
  active_customers = n_distinct(CustomerID[Subscription_status == "active"]),
  retention_rate = active_customers / total_customers
 )
```

📌 **Interpretation**

- Uses *latest subscription state only*
- This is the **CEO-friendly retention number**

2️⃣ Customer-Level Retention Status (Who stayed, who didn't)

```
customer_status <- subs %>%
 filter(Subscription_latest == 1) %>%
 select(CustomerID, Product_name, Offer_period, Subscription_status)

customer_status
```

📌 You can now answer:

- "Customer 101 is retained via annual winback"
- "Customer 104 is churned"

3️⃣ Cohort Retention (Gold Standard)

Step 1: Identify First Subscription (Cohort)

```
cohort <- subs %>%
  group_by(CustomerID) %>%
  summarise(cohort_month = min(start_month))
```

Step 2: Join Back & Calculate Months Since Start

```
cohort_data <- subs %>%
  left_join(cohort, by = "CustomerID") %>%
  mutate(
    months_since_start =
      interval(cohort_month, start_month) %/% months(1)
  )
```

Step 3: Cohort Retention Matrix

```
cohort_retention <- cohort_data %>%
  filter(Subscription_status == "active") %>%
  group_by(cohort_month, months_since_start) %>%
  summarise(
    retained_customers = n_distinct(CustomerID),
    .groups = "drop"
  )

cohort_retention
```

📌 Interpretation

- Month 0 = acquisition
- Month 1+ = retention
- This is what beIN / Netflix / DAZN use internally

4️⃣ Offer Period Retention (Monthly vs Annual)

```
offer_retention <- subs %>%
  filter(Subscription_latest == 1) %>%
  group_by(Offer_period) %>%
  summarise(
    customers = n_distinct(CustomerID),
    active = n_distinct(CustomerID[Subscription_status == "active"]),
    retention_rate = active / customers
  )

offer_retention
```

📌 **Business takeaway**

- Annual almost always shows **higher retention**
- Monthly drives churn but flexibility

5️⃣ Product Retention (Which product holds users?)

```
product_retention <- subs %>%
  filter(Subscription_latest == 1) %>%
  group_by(Product_name) %>%
  summarise(
    customers = n_distinct(CustomerID),
    retained = n_distinct(CustomerID[Subscription_status == "active"]),
    retention_rate = retained / customers
  )

product_retention
```

📌 Example insight:

> “4KTotal shows stronger retention than XYZTotal”

6️⃣ Continuous Renewal vs Break & Rejoin

This is **VERY important** for OTT.

```
renewal_behavior <- subs %>%
  group_by(CustomerID) %>%
  summarise(
    subscriptions = n_distinct(Subscription_ID),
    max_version = max(Subscription_version)
  ) %>%
  mutate(
    behavior = case_when(
      subscriptions == 1 & max_version > 1 ~ "continuous_renewal",
      subscriptions > 1 ~ "break_and_rejoin",
      TRUE ~ "single_term"
    )
  )

renewal_behavior
```

📌 Interpretation:

- **Continuous renewal** = healthy loyalty
- **Break & rejoin** = pricing / content issue

7️⃣ Winback Rate (Recovery Power)

```
winback <- subs %>%
  group_by(CustomerID) %>%
  summarise(
    churned = any(Subscription_type == "churned"),
    winback = any(Subscription_type == "winback")
  )

winback %>%
  summarise(
    churned_customers = sum(churned),
    winback_customers = sum(winback),
    winback_rate = winback_customers / churned_customers
  )
```

📌 **Business insight**

- High winback → good brand pull
- Low winback → permanent churn risk

8️⃣ Customer Lifetime (Tenure)

```
lifetime <- subs %>%
  group_by(CustomerID) %>%
  summarise(
    first_start = min(Subscription_start_date),
    last_expiry = max(Expiry_date),
    lifetime_days = as.numeric(last_expiry - first_start)
  )

lifetime
```

📌 Used for:

- LTV models
- Pricing decisions
- Promo effectiveness

9️⃣ Country-Aware Retention (VPN-safe)

Use **dominant country per customer**:

```
dominant_country <- subs %>%
  count(CustomerID, Country) %>%
  group_by(CustomerID) %>%
  slice_max(n, n = 1)

country_retention <- subs %>%
  filter(Subscription_latest == 1) %>%
  left_join(dominant_country, by = "CustomerID") %>%
  group_by(Country) %>%
  summarise(
    customers = n_distinct(CustomerID),
    retained = n_distinct(CustomerID[Subscription_status == "active"])
  )

country_retention
```

🔟 Executive Retention KPIs (From This Data)

KPI	Meaning
Logo Retention %	Active customers
Cohort Retention	Stickiness over time
Annual vs Monthly Retention	Contract efficiency
Product Retention	Content value
Continuous Renewal %	Loyalty
Winback Rate	Brand strength
Avg Lifetime (days)	Long-term value

🚀 What I Recommend Next

This is now **dashboard-ready**. Next logical steps:

- Add **usage minutes** → engagement-driven churn
- Build **cohort heatmap**
- Survival analysis (Kaplan–Meier)
- Predict churn using last 30-day inactivity

If you want, next I can:
👉 Turn this into **Power BI / Shiny dashboard logic**
👉 Add **visual cohort heatmap code**
👉 Combine **usage + subscription**

Just tell me 👍