
FalconAirlines.

PASSENGER SATISFACTION ANALYSIS & CUSTOMER CHURN COST ESTIMATION.

I. ABOUT THE DATA & LIMITATION.

Assumptions about the survey data:

1. **Integrity**

We assume customers are providing their honest and careful feedback for the survey.

2. **Situation & context**

We assume the customers aren't hurry to complete the survey after arriving at the destination and experiencing the service since most customers (about 69%) are business travellers as they could be busy with work. This would result in a bias outcome.

3. **Question & feedback**

We assume those survey questions are what the customers care about as the customers can't fill in any feedback other than choosing on a 1–5 scale based on those fixed questions. The outcome will not be holistic if the company assumes these are the only questions that matter to the customers. Additional research has to be done to obtain a more accurate picture of our customers' preference and feedback.

Limitation:

1. The features from the survey data is not suitable for making future prediction on the customers' satisfaction. The goal is to perform data mining to gain feature insight and estimate the customer churn cost.
2. Survey is not perfect method for gathering customers' preference and feedback, hence additional research need to be done to obtain a more holistic picture.

2. CUSTOMER CHURN COST.

Churn cost rough calculation:

The purpose of this formula is to **demo the analysis process in business context**. It does not follow the actual methods calculated by the airline companies.

Life time value

= (Luggage + Food + Class) * Two-way flight * Frequency per year.

Potential loss due to customer churn:

| Satisfaction <fctr> | CustomerType <fctr> | total_customer_life_time_value <chr> |
|-------------------------|------------------------|---|
| neutral_or dissatisfied | disloyal_customer | \$23,957,000 |
| neutral_or dissatisfied | loyal_customer | \$38,268,350 |
| neutral_or dissatisfied | NA | \$7,683,650 |
| satisfied | disloyal_customer | \$9,029,150 |
| satisfied | loyal_customer | \$93,829,650 |
| satisfied | NA | \$11,864,000 |

38.3M + 3.8M (half) = **42.1M**

| Satisfaction <fctr> | CustomerType <fctr> | TypeTravel <fctr> | total_customer_life_time_value <chr> |
|-------------------------|------------------------|----------------------|---|
| neutral_or dissatisfied | disloyal_customer | business_travel | \$22,931,550 |
| neutral_or dissatisfied | disloyal_customer | personal_travel | \$52,950 |
| neutral_or dissatisfied | disloyal_customer | NA | \$972,500 |
| neutral_or dissatisfied | loyal_customer | business_travel | \$29,359,650 |
| neutral_or dissatisfied | loyal_customer | personal_travel | \$6,861,000 |
| neutral_or dissatisfied | loyal_customer | NA | \$2,047,700 |
| neutral_or dissatisfied | NA | business_travel | \$6,808,650 |
| neutral_or dissatisfied | NA | personal_travel | \$875,000 |
| satisfied | disloyal_customer | business_travel | \$8,676,900 |
| satisfied | disloyal_customer | personal_travel | \$9,750 |

29.4M + 6.9M + 2M = **38.3M**

3.4M + 0.45M = **3.85M** (half)

`NA` refers to undefined customer / travel type.

3. MODEL COMPARISON.

Performance for each metrics:

| model <chr> | accuracy <dbl> | f_meas <dbl> | kap <dbl> | precision <dbl> | roc_auc <dbl> | sens <dbl> |
|-----------------------|--------------------------|------------------------|---------------------|---------------------------|-------------------------|----------------------|
| Model XGB | 0.9342963 | 0.9263206 | 0.8670375 | 0.9214636 | 0.9832798 | 0.9312292 |
| Model Random Forest | 0.9988678 | 0.9987242 | 0.9977065 | 0.9982611 | 0.9999931 | 0.9991877 |
| Model Decision Tree | 0.9180333 | 0.9071656 | 0.8337917 | 0.9114156 | 0.9651462 | 0.9029551 |
| Model GLM Logistic | 0.8329616 | 0.8111264 | 0.6614002 | 0.8135725 | 0.9067672 | 0.8086950 |

On TRAIN set.

| model <chr> | accuracy <dbl> | f_meas <dbl> | kap <dbl> | precision <dbl> | roc_auc <dbl> | sens <dbl> |
|-----------------------|--------------------------|------------------------|---------------------|---------------------------|-------------------------|----------------------|
| Model XGB | 0.9300337 | 0.9204872 | 0.8580281 | 0.9131143 | 0.9815786 | 0.9279801 |
| Model Random Forest | 0.9523523 | 0.9459152 | 0.9033418 | 0.9372460 | 0.9916267 | 0.9547461 |
| Model Decision Tree | 0.9154223 | 0.9028450 | 0.8279623 | 0.9052242 | 0.9661453 | 0.9004783 |
| Model GLM Logistic | 0.8324904 | 0.8094085 | 0.6600066 | 0.8038646 | 0.9057824 | 0.8150294 |

On TEST set (via stratified sampling).

4. MODEL INTERPRETATION.

Decision Tree:

| Feature | Importance |
|-------------------------------------|------------|
| "Inflight_entertainment" | "11735.61" |
| "Seat_comfort" | " 7535.76" |
| "Online_support" | " 5539.10" |
| "Ease_of_Onlinebooking" | " 4853.10" |
| "Online_boarding" | " 4023.79" |
| "Checkin_service" | " 2251.94" |
| "Food_drink" | " 1880.90" |
| "CustomerType_loyal_customer" | " 1140.57" |
| "Gate_location" | " 898.83" |
| "Departure_Arrival_time_convenient" | " 829.23" |
| "Cleanliness" | " 819.90" |
| "Onboard_service" | " 804.52" |
| "Inflightwifi_service" | " 646.38" |
| "Baggage_handling" | " 631.53" |
| "Class" | " 631.46" |
| "Leg_room_service" | " 613.96" |
| "TypeTravel_personal_travel" | " 410.03" |
| "Age_X_20_25_" | " 15.91" |
| "Age_X_65_above_" | " 10.32" |
| "DepartureDelayin_Mins" | " 6.97" |
| "Age_X_60_65_" | " 5.75" |
| "Age_X_35_40_" | " 4.45" |
| "Age_X_25_30_" | " 3.79" |
| "Age_X_45_50_" | " 1.71" |
| "Age_X_30_35_" | " 1.22" |

Decision Tree splitting path:

1. Inflight entertainment (≥ 5) > Loyal customer > Class ($= 3$) > Business travel > Check-In service (≥ 4)
2. Inflight entertainment (≥ 5) > Loyal customer > Class ($= 3$) > Business travel > Check-In service (≥ 4) > Seat comfort (≤ 4)
3. Inflight entertainment (≥ 5) > Loyal customer > Class ($= 3$) > Business travel > Check-In service (≥ 4) > Seat comfort (≥ 5) > Gate location (≥ 5)
4. Inflight entertainment (≥ 5) > Loyal customer > Class ($= 3$) > Personal travel > Leg room service (≥ 5)
5. Inflight entertainment (≥ 5) > Loyal customer > Class (≤ 2) > Seat comfort ($= 6$) > Check-In service (≥ 4)
6. Inflight entertainment (≥ 5) > Loyal customer > Class (≤ 2) > Seat comfort (≤ 5) > Departure arrival time convenient (≤ 5) > Food drink (≤ 4)
7. Inflight entertainment (≥ 5) > Loyal customer > Class (≤ 2) > Seat comfort (≤ 5) > Departure arrival time convenient (≤ 5) > Food drink (≥ 4) > Gate location (≥ 5)
8. Inflight entertainment (≥ 5) > Loyal customer > Class (≤ 2) > Seat comfort (≤ 5) > Departure arrival time convenient (≥ 5) > Personal travel
9. Inflight entertainment (≥ 5) > Loyal customer > Class (≤ 2) > Seat comfort (≤ 5) > Departure arrival time convenient (≥ 5) > Business travel > Seat comfort (≥ 5)

Logistic Regression:

| Feature | Log Odds | Odds Ratio |
|-------------------------------------|------------|------------|
| "(Intercept)" | "-11.5219" | "0.000" |
| "CustomerType_loyal_customer" | " 2.0452" | "7.731" |
| "TypeTravel_personal_travel" | " -0.7768" | "0.460" |
| "Inflight_entertainment" | " 0.7158" | "2.046" |
| "Age_X_20_25_" | " 0.6252" | "1.869" |
| "Class" | " 0.3470" | "1.415" |
| "Age_X_65_above_" | " -0.3296" | "0.719" |
| "Onboard_service" | " 0.3095" | "1.363" |
| "Checkin_service" | " 0.2986" | "1.348" |
| "Seat_comfort" | " 0.2916" | "1.339" |
| "Leg_room_service" | " 0.2462" | "1.279" |
| "Ease_of_Onlinebooking" | " 0.2250" | "1.252" |
| "Departure_Arrival_time_convenient" | " -0.2233" | "0.800" |
| "Food_drink" | " -0.1926" | "0.825" |
| "DepartureDelayin_Mins" | " -0.1810" | "0.834" |
| "Online_boarding" | " 0.1587" | "1.172" |
| "Gate_location" | " 0.1285" | "1.137" |
| "Online_support" | " 0.1010" | "1.106" |
| "Baggage_handling" | " 0.0934" | "1.098" |
| "Cleanliness" | " 0.0917" | "1.096" |
| "Inflightwifi_service" | " -0.0651" | "0.937" |

5. FIND OPTIMAL DISCRIMINANT THRESHOLD.

Reducing the cost of false negative:

To avoid underestimating (due to false negative) the customer churn cost for loyal business and personal travellers who are neutral/dissatisfied, the discriminant threshold should at least be around **0.75**.

| Satisfaction <fctr> | CustomerType_loyal_customer <dbl> | TypeTravel_personal_travel <dbl> | total_customer_life_time_value <chr> |
|------------------------|--------------------------------------|-------------------------------------|---|
| 0 | 0 | 0 | \$20,283,150 |
| 0 | 0 | 1 | \$43,000 |
| 0 | 1 | 0 | \$26,217,300 |
| 0 | 1 | 1 | \$5,949,150 |
| 1 | 0 | 0 | \$7,419,450 |
| 1 | 0 | 1 | \$6,500 |
| 1 | 1 | 0 | \$73,082,100 |
| 1 | 1 | 1 | \$5,140,050 |

Actual customer life time value on TRAIN with **0.5** threshold.

| .pred_class <fctr> | CustomerType_loyal_customer <dbl> | TypeTravel_personal_travel <dbl> | total_customer_life_time_value <chr> |
|-----------------------|--------------------------------------|-------------------------------------|---|
| 0 | 0 | 0 | \$23,351,850 |
| 0 | 0 | 1 | \$44,300 |
| 0 | 1 | 0 | \$26,604,450 |
| 0 | 1 | 1 | \$6,496,500 |
| 1 | 0 | 0 | \$4,350,750 |
| 1 | 0 | 1 | \$5,200 |
| 1 | 1 | 0 | \$72,694,950 |
| 1 | 1 | 1 | \$4,592,700 |

Actual customer life time value on after adjusting the threshold to **0.75**.

6. ESTIMATED REDUCTION IN CUSTOMER CHURN COST.

Reduction in churn cost after improving the top 5 important features:

| Satisfaction <fctr> | CustomerType_loyal_customer <dbl> | TypeTravel_personal_travel <dbl> | total_customer_life_time_value <chr> |
|------------------------|--------------------------------------|-------------------------------------|---|
| 0 | 0 | 0 | \$2,810,450 |
| 0 | 0 | 1 | \$12,550 |
| 0 | 1 | 0 | \$3,643,400 |
| 0 | 1 | 1 | \$2,550,400 |
| 1 | 0 | 0 | \$1,088,650 |
| 1 | 0 | 1 | \$4,550 |
| 1 | 1 | 0 | \$10,460,350 |
| 1 | 1 | 1 | \$2,280,450 |

Total 6.19M

Customer churn cost on TEST set (before improving the rating)

| .pred_class <fctr> | CustomerType_loyal_customer <dbl> | TypeTravel_personal_travel <dbl> | total_customer_life_time_value <chr> |
|-----------------------|--------------------------------------|-------------------------------------|---|
| 0 | 0 | 0 | \$3,468,550 |
| 0 | 0 | 1 | \$14,500 |
| 0 | 1 | 0 | \$1,332,050 |
| 0 | 1 | 1 | \$1,320,450 |
| 1 | 0 | 0 | \$430,550 |
| 1 | 0 | 1 | \$2,600 |
| 1 | 1 | 0 | \$12,771,700 |
| 1 | 1 | 1 | \$3,510,400 |

Total 2.65M

3.54M of reduction (57%)

Customer churn cost on TEST set (after improving the rating)

Reduction in neutral/dissatisfied proportion:

| | percentile | prob |
|-------|------------|----------|
| [1,] | "0%" | "0.0000" |
| [2,] | "2.5%" | "0.0000" |
| [3,] | "5%" | "0.2941" |
| [4,] | "7.5%" | "0.3012" |
| [5,] | "10%" | "0.3012" |
| [6,] | "12.5%" | "0.3012" |
| [7,] | "15%" | "0.3012" |
| [8,] | "17.5%" | "0.3237" |
| [9,] | "20%" | "0.3237" |
| [10,] | "22.5%" | "0.6031" |
| [11,] | "25%" | "0.6031" |
| [12,] | "27.5%" | "0.6031" |
| [13,] | "30%" | "0.7554" |

| neutral_or dissatisfied | satisfied |
|-------------------------|-----------|
| 0.4526766 | 0.5473234 |

15–22.5% of reduction between **0.5–0.75**
discriminant threshold

7. RECOMMENDATION & NEXT STEPS.

Recommendation/Next steps on neutral/dissatisfied customers & user research:

1. Extend the user research to interviews with different segment of customers after performing the customer segmentation.
2. Extend the user research to value mapping to map out and analyse the customer touchpoint and user journey.
3. Investigate the `Departure_Arrival_time_convenient` and `Food_drink` feature as there could be a data error.
4. Perform feature prioritisation by analysing the business capabilities and feasibility for the top N important features, and the estimated reduction in customer churn cost.
5. Pay extra attention to loyal business travellers since the customer churn cost is higher as they travel more frequently than the personal travellers.

THE END.