Baggage Unloading Process in Airline Operations

Executive Summary

This report evaluates efficiency of baggage unloading using flight-level data, with **a focus on the probability that a flight's baggage delivery time is under 21 minutes.** This threshold is important as it is tied to customer service expectations and delivery time guarantees. Of 99,174 records, 14.6% of rows contained data issues, primarily full duplicates. 8,484 rows were removed, leaving 90,690 for analysis.

Visual analysis suggested both normal and log-normal distributions could model the data. While the raw data was less skewed, the log normal model was chosen due to its alignment with nature of real-world baggage delivery times.

Empirical probability of deliveries completed within 21 minutes is 81.62%, while a lognormal model distribution estimates a lower 80.14% probability. Despite being numerically small, this difference is statistically significant as confirmed by a one-sample hypothesis test and a 95% confidence interval.

Both approaches have strengths and limitations, highlighting the need to balance model assumptions and real-world data to make actionable interferences.

Data Pre-Processing

All time fields are stored as timestamps with minute level granularity, therefore there is no need for rounding or aggregation. They can directly be parsed for computing time differences.

The following issues were identified after performing logical checks during exploratory analysis (*See Exhibit A for full list of data quality checks performed*):

1. Existing Issues/Discrepancies and their pervasiveness

Issue	Description	Pervasiveness (% of rows)
Complete Duplicates	Identical rows across all columns	14.385%
Partial Duplicates	Same flight number and arrival times	0.022%
Origin = Dest	Identical origin and destination airport codes	0.035%
Bag Drop Times > 0 & Expected Bags = 0	BagDropTime is present but ExpectedBags is missing	0.082%
Bag Drop Times = 0 & Expected Bags > 0	ExpectedBags > 0 but no recorded BagDropTime	0.118%

Figure 1: Data Issues and pervasiveness

The number of rows with at least one data quality issue is 14,451, ~14.571% of all rows (Not sum of rows with each issue as some rows have multiple). The most common is complete duplicates while others occur less frequently. This poses a risk to analysis as it can skew metrics and inflate counts.

2. Addressing Issues via Statistical Tools and Software

Data processing for issues addressed using Python within a Jupyter Notebook environment (Exhibit B.2).

A total of 8484 rows were removed, reducing the rows in the dataset to 90,690.

Issue	Action Taken	Justification
Complete Duplicates	Removed	Full duplicates inflate count and delivery time
Partial Duplicates	Kept	Delivery times exist; assumed real flights with incorrect meta data (data entry issue)
Origin = Dest	Kept	Delivery times exist; assumed real flights with routing data entry error
Bag Drop Times > 0 & Expected Bags = 0	Kept	Delivery times exist; assumed real flights with incorrect bag count data entry error
Bag Drop Times = 0 & Expected Bags > 0	Removed	Lack of key outcome (delivery time data), biases average time downwards, limiting analytical value

Figure 2: Action taken on Data Issues and reasoning

Data Analysis

3. Analysis of "Baggage Delivery Time" Variable

$$DeliveryDurationMins = LastBagDropTime - ActualArrival$$ (Exhibit B.1)

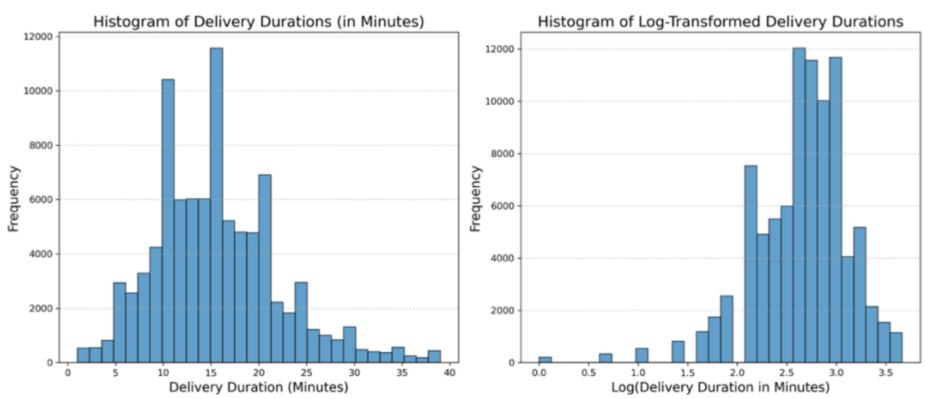

Figure 3: Histogram Plots of DeliveryDurationMins + Log Transformed Plot

The distribution of baggage delivery times is bell-shaped with slight right skew, suggesting it is log normal. Applying a log transformation indeed reduces right skew but introduces left skew. Importantly, the cleaned dataset contains no records where DeliveryDurationMins = 0, ensuring all rows can be log-transformed without encountering undefined values, enabling seamless log transformation while preserving all rows.

	Raw	Log-Transformed
Count	90690	90690
Mean	15.549939	2.647437
95% Confidence Interval	(15.508, 15.592)	-
Standard Deviation	6.504967	0.468989
Min	1.000000	0.000000
Max	39.000000	3.663562
Skewness	0.7424530536482606	-1.0492929348068154

Figure 4: Summary Statistics of Raw and Log-Transformed DeliveryDurationMins

4. Probability that Baggage Delivery for random flight < 21 mins

4.1 Empirical Approach

The empirical probability was calculated using Excel functions =COUNTIF and =COUNT (Exhibit C).

$$P(Delivery < 21) = \frac{Count\ Flights < 21\ Mins\ Delivery}{Count\ Total\ Flights} = \frac{74023}{90690} \approx 0.8162\ or\ 81.62\%$$

4.2 Theoretical Approach via Log Normal Modelling

Both raw and log-transformed delivery durations appear bell-shaped with varying skewedness (see Figure 3). While the absolute skewness the raw distribution is lower and suggests it is a more normal distribution, a log normal model was ultimately chosen to compute the probability of interest (see Figure 4 for skewness). The reason is driven by real-world context, where baggage delivery times peak earlier, and have a tail of rarer extreme delays. This calculation was validated using Excel function =NORM.DIST (Exhibit G).

$$X\ (Log\ Point\ of\ Interest - 21\ mins\ ) = \ln(21) = 3.045$$

$$\mu\ (mean) = 2.647437 \qquad\qquad \sigma\ (stdev) = 0.468989$$

$$z = \frac{X - \mu}{\sigma} = \frac{3.045 - 2.647437}{0.468989} \approx 0.847$$

$$P(Delivery < 21) \approx 80.14\%$$

Using a Z-Score of 0.847, the theoretical probability is estimated to be 80.14%

5. Empirical vs Theoretical Methods

Approach	P (Delivery < 21)	95% CI	Difference Vs .Empirical	Notes
Empirical	81.62%	(81.4%, 81.9%)	-	Based on 90,690 flights; “success” = delivery < 21 mins
Log-Normal Model	80.14%	-	-1.48%	Based on log-normal distribution
Normal Model (Not Selected)	*79.89%*	-	*-1.73%*	*Based on normal distribution (Exhibit D)*

Figure 5: Probability comparisons

The Empirical and Theoretical approaches yielded different outcomes, with a difference of 1.48%. While the empirical probability suggests an 81.62% probability a flight has a <21 minutes baggage delivery time, the log-normal model estimates a lower 80.14%. This value falls outside of the 95% confidence interval of the empirical proportion meeting success criteria of <21 minutes delivery (81.4%, 81.9%) suggesting a statistically significant different between empirical and model probabilities (Exhibit E on proportion CI).

A one-sample hypothesis test for proportions at a 5% significance level confirms this, rejecting the null hypothesis the empirical and theoretical probabilities are equal, concluding that there is a statistically significant difference between the two approaches. This highlights that strengths and limitations must be considered which method to use (Exhibit F on Hypothesis Test).

	Advantages	Disadvantages
Empirical	Reflects true behavior of dataset Good for irregular data, works even if data is not normally distributed or contains outliers Simple to calculate and interpret	Only analyzes what is already in dataset with limited extrapolation capability beyond sample Sensitive to extreme values such as outliers and errors The calculation process does not allow for understanding of distribution, making it limited for understanding patterns
Theoretical	Enables prediction beyond samples if variable of interest behaves closely to chosen model Allows estimation of tail probabilities, which are theoretically unlimited instead of just with observed values	May overlook real-world anomalies, systematic issues, and edge cases Relies on choosing appropriate distribution to model data on, otherwise results can be mis interpretive Model estimates may deviate from empirical data

Summary & Reflection

The analysis highlights the need to validate models against observed outcomes beyond the initial differences. A 1.48% may seem numerically minor, but it was a statistically significant difference between model performance and observed outcome. Nevertheless, both models agree that most flights meet the 21 minutes baggage delivery threshold.

A challenge involved choosing the model distribution for the theoretical approach, as it required deciding between using summary statistics or assumptions on real world process. Whereas the raw data was less skewed, a log-transformation is more aligned to real world behavior.

Data quality also posed challenges. Decisions on whether to retain or discard rows were made without full context. For example, records origin=destination airport codes were kept if they had usable delivery times — which preserves aggregate analysis but risks introducing bias in location-specific interpretations.

Lastly, a key consideration is that the cleaned dataset did not have rows with Delivery Duration equals to zero. This was coincidental and allowed a smooth log transformation, but future datasets may have such values. If present, they would need to be adjusted or excluded, complicating analysis and potentially weakening summary insights.

Exhibits

A. Existing Data Issues Checked

Logical checks were performed on raw dataset, with the issues highlighted red failing:

1. FirstBagDropTIme >= ActualArrival
2. ExpectedBagsCount >= 0
3. No null values
4. Exact duplicate rows
5. Partial duplicate rows (Same flight number and arrival time)
6. Origin != Destination
7. Expected Bags = 0, Bag Drop Times > 0
8. Expected Bags >= 0, Bag Drop Times = 0

```
# Sample Check for Exact Duplicates
check6 = df[df.duplicated(keep=False)]
audit_results['Exact Duplicate Rows'] = len(check6)
bad_rows['Exact Duplicate Rows'] = check6
```

B. Data Processing

B.1 DeliveryDurationMins Calculation & Code Snippet

```
# Compute Delivery Time in minutes.
df['DeliveryDurationMins'] =
(df['LastBagDropTime'] - df['ActualArrival']).dt.total_seconds() / 60
```

B.2 Data Cleaning Code Snippet

```
# Remove duplicates and remove rows with bag count but no wait times
df_cleaned = df.drop_duplicates()
df_cleaned = df_cleaned[~((df_cleaned['ExpectedBagsCount'] > 0) &
(df_cleaned['DeliveryDurationMins'] == 0))]
```

C. Empirical Probability

```
# DeliveryDurationMins in Column H:
= COUNTIF(H2:H90691, "<21") / COUNT(H2:H90691)
= COUNTIF(90690, "<21") / COUNT(90690)
```

D. Theoretical Probability – Normal Distribution (Not Selected)

$$X\ (Point\ of\ Interest - 21\ mins\ ) = 21$$

$$\mu\ (mean) = \ 15.549939 \qquad\qquad \sigma\ (stdev) = 6.504967$$

$$z = \frac{X - \mu}{\sigma} = \frac{21 - 15.549939}{6.504967} \approx \ 0.8378$$

$$P(Delivery < 21) \approx 79.89\%$$

E. Confidence Interval for Proportion

$p = 0.8162$ (empirical success proportion of < 21 mins delivery)

$z_{a/2} = 1.96$ (z-score for 95% CI)

$n = 90690$ (sample size)

$$CI = p \pm z_{a/2}\sqrt{\frac{p(1-p)}{n}} = 0.8162 \pm 1.96\sqrt{\frac{0.8162(1-0.8162)}{90690}} = 0.8162 \pm 0.00242$$

F. One-Sample Hypothesis Test for Proportion

Empirical probability, E, is 81.62%, while Theoretical, T, is 80.14%

Null Hypothesis $H_0: E = T$

Alternative Hypothesis $H_1: E \neq T$

$p = 0.8162$ (observed proportion)

π_0 = 0.8014 (hypothesized proportion from model)

$n = 90690$ (sample size)

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.8162 - 0.8014}{\sqrt{\frac{0.8014(1-0.8014)}{90690}}} \approx 10.79$$

p-value (Two-Tailed) ≈ 0.000< 0.05

Conclusion: Reject null hypothesis at 5% significance level. E and T statistically different.

G. Theoretical Probability Calculation in Excel

```
# Log Transform Raw Data from Column H into Column G.
= LN(H2)
# Apply down column

# Compute Theoretical Probabilities
= NORM.DIST(LN(21), AVERAGE(G2:G100000), STDEV.S(G2:G100000, 1)
```