

Duale Hochschule Baden-Württemberg Mannheim

Gruppenarbeit

Ready Jet Go

Studiengang Wirtschaftsinformatik

Studienrichtung Data Science

Verfasser:

Eric Echtermeyer, Lasse Friedrich,

Ahmet Ilgar Korkmaz, Benedikt Prisett, David Schäfer

Matrikelnummer:

6373947, 9924680, 6357368 5709658, 7086451

Kurs:

WWI-21-DSA

Studiengangsleiter:

Prof. Dr.-Ing. habil. Dennis Pfisterer

Dozent:

Janett Betz und Enzo Hilzinger

Bearbeitungszeitraum:

08.05.2024 – 17.07.2024

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
1 Einleitung	1
2 Datenvorverarbeitung	2
3 Explorative Datenanalyse	3
3.1 Geographische Visualisierung	3
3.2 Weiterführende Visualisierungen	7
4 Gewichtswertanalyse	10
4.1 Flugzeugtypen-Clustering	11
4.2 Grundlagenanalyse	13
4.2.1 Identifizierung relevanter Daten	13
4.2.2 Datenverständnis aufbauen	14
4.3 Schätzwertoptimierung	15
5 Prozessanalyse	20
5.1 Process-Mining	20
5.2 Dashboards	21
5.3 Auswertung	24
6 Schlussbetrachtung	27
6.1 Lessons Learned und Reflexion	27
6.2 Erfüllte und nicht erfüllte Kriterien	28
6.3 Ausblick	29
Anhang	
A Anhang Gewichtswertanalyse	31
B Anhang Prozessanalyse	33
B.1 Prozessablauf	34
B.1.1 Airline ABCD	34

Inhaltsverzeichnis

B.1.2 Airline MNOP	36
B.1.3 Airline ZYXW	38
B.2 Dashboards	39
Literaturverzeichnis	43

Abbildungsverzeichnis

3.1	Landkarte aller Ankunftsflughäfen	5
3.2	Landkarte der extrahierten Routen	6
3.3	Landkarte der extrahierten Zwischenstopps	7
3.4	Dashboard mit weiteren Visualisierungen	9
4.1	Größe der Flugzeugflotte (blau) sowie Anzahl durchgeföhrter Flüge (grün) und existierender Versionen (orange) pro Untertyp	11
4.2	Reichweite in km und Gewicht in kg sowie Anzahl durchgeföhrter Flüge (Größe) pro Untertyp	13
4.3	Visualisierung der Beziehung grundlegender Gewichtsdaten	15
4.4	Mittelwert (Linie) und Standardabweichung (Fläche) der Differenz aus erwartetem und tatsächlichem ZFW im Zeitverlauf, aggregiert über alle Flüge	16
4.5	Durchschnittliche Abweichung des tatsächlichen TOW von den Vorhersagen des neuronalen Netzes sowie der Schätzungen im Datensatz für die Flüge der Fluggesellschaft AB in den Testdaten	18
4.6	Durchschnittliche Abweichung des tatsächlichen TOW von den Vorhersagen des neuronalen Netzes sowie der Schätzungen im Datensatz für die Flüge der Fluggesellschaft ZY in den Testdaten	18
5.1	Übersicht der sechs erstellen Dashboards. Detailansichten sind im Anhang B.2 zu finden.	24
A.1	Anzahl der eingekochten Passagiere und geschätztes Passagiergewicht	31
A.2	Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft AB	31
A.3	Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft MN	32
A.4	Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft ZY	32
B.1	Der häufigste Prozessablauf der Fluggesellschaft ABCD (42,4% der Aktivitäten)	34

Abbildungsverzeichnis

B.2	Prozessablauf der Fluggesellschaft ABCD mit erhöhten Detailgrad (67,4% der Aktivitäten)	35
B.3	Der häufigste Prozessablauf der Fluggesellschaft MNOP (39,9% der Aktivitäten)	36
B.4	Prozessablauf der Fluggesellschaft MNOP mit erhöhten Detailgrad (45,4% der Aktivitäten)	37
B.5	Der häufigste Prozessablauf der Fluggesellschaft ZYXW (85,5% der Aktivitäten)	38
B.6	Prozessablauf der Fluggesellschaft ZYXW mit erhöhten Detailgrad (97,1% der Aktivitäten)	39
B.7	Dashboard „Data Overview“ mit einem allgemein Überblick der relevanten KPIs.	40
B.8	Dashboard „Process Flow“ mit einer Darstellung des Prozessdiagramms.	40
B.9	Dashboard „Process Time“ mit Einblicken in die Durchlaufzeiten.	41
B.10	Dashboard „Airport Drilldown“ mit Vergleich der Flughäfen.	41
B.11	Dashboard „Process Automation“ mit Darstellungen des Automatisierungsgrads.	42
B.12	Dashboard „Variant Explorer“ mit Darstellung von Prozess Varianten.	42

Abstract

Diese Studie analysiert Logdaten des Flugzeugbeladungsprozesses von drei Tochtergesellschaften der Lufthansa AG mittels Process Mining und Machine Learning. Die Hauptziele waren (1) die Analyse des Beladeprozesses verschiedener Fluggesellschaften, (2) die Erstellung aussagekräftiger Datenvisualisierungen und (3) die Entwicklung eines Vorhersagemodells zur Optimierung der Flugzeuggewichtsplanung. Die Untersuchung umfasste eine Prozessanalyse, explorative Datenanalyse und Gewichtswertoptimierung. Die Ergebnisse zeigen, dass die vorliegenden Daten Inkonsistenzen und Qualitätsmangel aufweisen, was die Analyse erschwert. Es konnte kein einheitlicher Standardprozess für die Flugzeugbeladung identifiziert werden. Stattdessen variieren die Prozessabläufe sowohl zwischen den Airlines als auch zwischen einzelnen Flughäfen erheblich. Trotz dieser Variabilität zeigen sich signifikante Effizienzunterschiede zwischen den untersuchten Fluggesellschaften. Insbesondere eine Airline wies mit durchschnittlich 17 Aktivitäten pro Flugplanung einen deutlich schlankeren Prozess auf als die anderen mit 132 bzw. 267 Aktivitäten. Die entwickelten Dashboards ermöglichen dem Auftraggeber eine kontinuierliche Prozessüberwachung und -optimierung. Die Studie unterstreicht die Notwendigkeit verbesserter Datenkonsistenz und Prozessstandardisierung im Flugzeugbeladungsmanagement und bietet Ansatzpunkte für zukünftige Optimierungen zur Steigerung der betrieblichen Effizienz.

1 Einleitung

Die effektive Datennutzung ist in der heutigen digitalisierten Wirtschaft von zentraler Bedeutung für den Unternehmenserfolg. Eine Studie von Statista (2023) zeigt jedoch, dass 66% der befragten Unternehmen das Potenzial ihrer Daten nur unzureichend aus schöpfen [2]. Diese Problematik betrifft auch die Lufthansa AG, Deutschlands größte Fluggesellschaft [1].

Das Projekt „Ready Jet Go“ zielt darauf ab, einen Ausschnitt der Logdaten von drei Tochtergesellschaften der Lufthansa AG umfassend zu analysieren. Der Fokus liegt dabei auf drei Hauptaspekten: (1) Analyse des Beladeprozesses verschiedener Fluggesellschaften, (2) Erstellung aussagekräftiger Datenvisualisierungen und (3) Entwicklung eines Vorhersagemodells zur Optimierung der Flugzeuggewichtsplanung.

Die Projektdurchführung erfolgt in drei Phasen. In der Planungsphase werden nach dem Projekt-Kick-Off ein Projektstrukturplan und ein Lastenheft erstellt. Diese Dokumente bilden die Grundlage für die weitere Projektarbeit. In der anschließenden Implementierungsphase werden die definierten Anforderungen umgesetzt. Hierbei kommen iterative, an SCRUM angelehnte Methoden zum Einsatz, die durch regelmäßige Teamtreffen und Aufgabenverteilung gekennzeichnet sind. In der abschließenden Kommunikationsphase werden die Ergebnisse in einem Abschlussbericht dokumentiert und dem Auftraggeber präsentiert.

Dieses Projekt stellt einen ersten Schritt zur verbesserten Datennutzung innerhalb der Lufthansa-Gruppe dar. Es legt die Grundlage für datengestützte Entscheidungsprozesse und trägt somit potenziell zur Steigerung der betrieblichen Effizienz bei. Die Kombination aus umfassender Datenanalyse, Visualisierung und prädiktiver Modellierung verspricht wertvolle Einblicke in betriebliche Abläufe und Optimierungspotenziale. Die gewonnenen Erkenntnisse können als Basis für zukünftige strategische Entscheidungen und operative Verbesserungen dienen.

2 Datenvorverarbeitung

Grundlage sämtlicher Projektziele ist eine umfassende und qualitativ hochwertige Datenvorverarbeitung. Bei den von den Projektauftraggebern zur Verfügung gestellten Daten handelt es sich um Logging-Dateien, welche für jede der Flugairlines separat als CSV-Datei abgespeichert wurden. Da diese Dateien in ihrer Struktur ähnlich sind, wurde ein dynamischer DataLoader mithilfe von Python entwickelt. Dieser mehrere Komponenten, welche untergliedert werden können in *CSV_Cleaner*, *Preprocessor*, *DF_Cleaner* und *Regex_Extractor*. Während der *CSV_Cleaner* Fehler in den übergebenen CSV-Dateien dynamisch repariert, orchestriert der *Preprocessor* alle weiteren Schritte der Datenvorverarbeitung. Zu Beginn des *Preprocessors* werden grundlegende Verarbeitungsschritte absolviert, wie beispielsweise die Extraktion der Header-Kategorie oder das Ableiten eines geeigneten Identifikators für einzelne Flüge. Hierbei wurden Flüge nach dem kombinierten Schlüssel von Airline-Code, Flugnummer, Flugzeitstempel und Abflugflughafen gruppiert. Anschließend erstellt der *Preprocessor* iterativ mit dem *Regex_Extraktor* neue Tabellen mit den extrahierten Daten für unterschiedliche Aktionsarten wie *CalculateWeightAndTrimAction*. Um die Größe der einzelnen Tabellen minimal zu halten, wurden diese extrahierten Werte in separaten Dateien gespeichert, mit Fremdschlüsseln referenzierend auf den zugehörigen Flug. Da es öfters zu fehlerhaft dargestellten Werten innerhalb der Daten kam, da die zur Verfügung gestellten Daten falsch waren, mussten anschließend stets einzelne Attribute bereinigt werden. Hierfür wurde *DF_Cleaner* verwendet. Dieser iterative Prozess zwischen *Regex_Extraktor* und *DF_Cleaner* wurde anschließend verwendet, um mehr als 41% sämtlicher Logging-Daten zu extrahieren, sodass 201 neue Attribute für jede Airline an zusätzlichen Daten zur Verfügung stehen. Als finales Resultat konnten zudem sämtliche Gewichts- und Betankungsdaten extrahiert werden, was essenziell für folgende Schritte der Datenanalyse ist. Während der Entwicklung der zuvor beschriebenen Komponenten zur Datenvorverarbeitung wurde stets auf eine möglichst hohe Wiederverwendbarkeit des Codes geachtet, um den zukünftigen Gebrauch dessen zu fördern. Daher stellt die gesamte erarbeitete Datenvorverarbeitungspipeline bereits das erste gelieferte Produkt dar, welches von den Auftraggebern zukünftig weiterhin verwendet werden kann.

3 Explorative Datenanalyse

Basierend auf ersten Erkenntnissen aus der Datenvorverarbeitung konnten verschiedene Flughäfen identifiziert werden. In diesem Zusammenhang wurden im Projektauftrag zwei Anforderungen definiert, welche ergänzende Analysen darstellen sollen, um einen genaueren Einblick in die Daten zu erreichen. Ziel in Anforderung 1 war es dabei vor allem geographische Daten zu analysieren und zu visualisieren. Durch das Vorliegen der Flughäfen konnte die genaue Zuordnung der geographischen Lage extrahiert und angezeigt werden. Ergänzend dient die Anforderung 2, welche weiterführende Analysen und Visualisierungen abdecken soll. Hierbei werden weitere Variablen analysiert und im nächsten Schritt visualisiert.

3.1 Geographische Visualisierung

Für die Bearbeitung der Aufgabe wurde zunächst das Tool Power BI von Microsoft verwendet. Bei der Bearbeitung der Aufgabe konnten jedoch die gewünschten Darstellungen mit Power BI nur teilweise erstellt werden. Bereits im Lasten- und Pflichtenheft wurde als Alternative Tableau genannt, welche im Anschluss für die Bearbeitung der Anforderung genutzt wurde.

In beiden Anforderungen ist somit das Laden der Daten in Tableau nötig. Hierbei kam es zu ersten Problemen, da die verarbeiteten CSV-Dateien zu groß und zu komplex sind, um sie einlesen zu können. Als Lösung wurden die Dateien zur Excel-Datei verarbeitet. Durch diesen Schritt wurde die Größe der Dateien von etwa 1 Gigabyte pro Datei auf etwa 50 Megabyte verringert, sodass sie ohne weitere Probleme eingelesen werden konnten.

Als Ergebnis der ersten Anforderung werden drei Landkarten mit verschiedenen Visualisierungen erstellt. Die Darstellungen wurden anschließend über die „Tableau Public“-Funktion in einer Website gehostet. Über den folgenden Link sind die Visualisierungen

einzusehen: <https://public.tableau.com/app/profile/ilgar.korkmaz/viz/ReadyJetGo-Visualisierungen/AlleAnkunftsflughafen?publish=yes>

Die erste Visualisierung stellt alle vorliegenden Abflug-Flughäfen dar. Diese Daten stammen dabei aus der bereits vorliegenden Spalte „departure_airport“. In dieser Spalte werden die Flughäfen nach ihren IATA-Codes gespeichert. Mit diesem internationalen Kürzel ist es möglich, die Koordinaten der Flughäfen mit Längen- und Breitengrad zu extrahieren. Anschließend konnten beide Datenquellen verknüpft werden, sodass eine Darstellung innerhalb einer Landkarte möglich ist. Neben dem Standort des Flughafens wurde hierbei noch nach dem Airline Code farbig getrennt. Neben dem Airline Code wird zudem noch die Häufigkeit der Flughäfen in den Flügen gezählt und anhand der Größe des Punktes dargestellt. Zur Kontrolle wurde zudem die gesamt Häufigkeitszahl mit den Ergebnissen von Kapitel 5 verglichen, wobei die Übereinstimmung die Korrektheit der Werte beweisen konnte.

In der ersten Visualisierung ist auffällig, dass sich die Flughäfen in 3 Regionen befinden: Europa, Amerika und Asien. Hierbei sind die Airline Codes auch größtenteils immer in einer Region vorhanden. Airline „AB“ nutzt dabei Flughäfen in Asien, Airline „MN“ in Europa und Nordamerika und Airline „ZX“ hauptsächlich in Südamerika.,„ZX“ nutzt dabei zudem noch einige Flughäfen, die in Europa und Nordamerika liegen. Innerhalb von Europa befinden sich die Flughäfen eher im Westen. Zusätzlich ist zu erkennen, dass die meisten Flughäfen in Südamerika liegen.

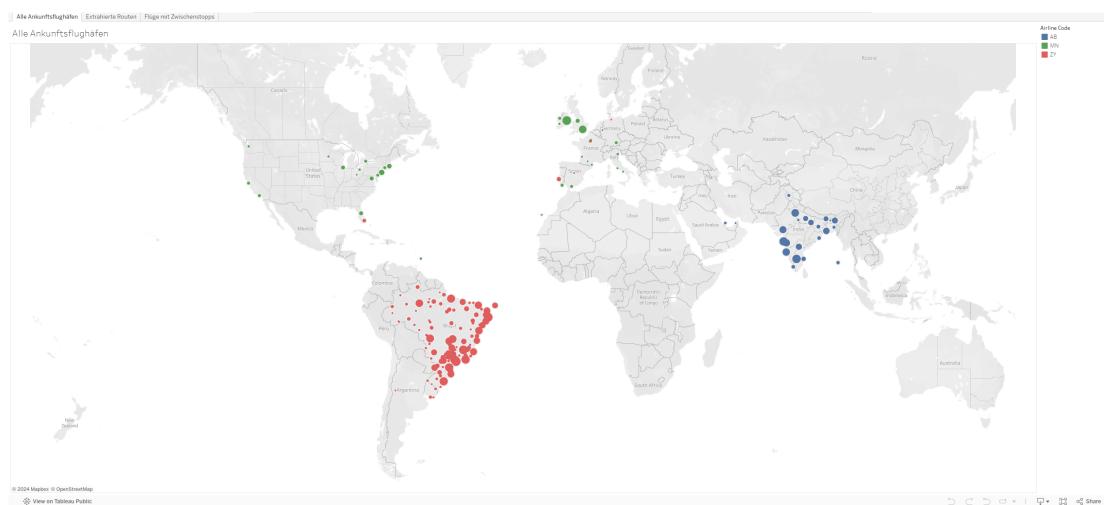


Abbildung 3.1: Landkarte aller Ankunftsflughäfen

Für die beiden anderen Visualisierungen wurden Daten aus der Spalte „entry_details“ extrahiert. Aus der Spalte wurden weitere Flughafen-Daten entdeckt, darunter der Abflug- und Ankunft-Flughafen. Diese wurden für die zweite Visualisierung verwendet, um Flugrouten anzuzeigen, mit dem Ziel ein besseres Verständnis über die Flugverläufe zu erlangen.

In diesem Schritt mussten erneut die Längen- und Breitengrade der Flughäfen basierend auf den IATA-Codes extrahiert werden. Zudem wurde ein zusätzlicher Schritt durchgeführt, um nicht nur die Flughäfen in der Landkarte anzuzeigen, sondern auch die Routen einzeichnen zu können. Hierbei wurde eine neue Spalte hinzugefügt, welche die gewünschte Route speichert. Diese verbindet den Abflug- und Ankunft-Flughafen mit einem Unterstrich (Bsp. BOM_DOH). Mit dieser neuen Spalte ist es im Anschluss möglich, die zwei Flughäfen mit einer Linie zu verbinden. In dieser Visualisierung wurde zudem eine transparente Linie gewählt, sodass die Intensität der Linie die Häufigkeit der Route anzeigt.



Abbildung 3.2: Landkarte der extrahierten Routen

In Abbildung 3.2 ist dabei zu erkennen, dass sowohl nationale als auch internationale Flüge vorliegen. In Nordamerika finden zum Beispiel keine nationalen Flüge statt, sondern größtenteils nur Flüge nach Europa. Generell ist in der Karte auch ersichtlich, dass internationale Flüge weniger stattfinden, als nationale Flüge, welches durch die Transparenz der Linien zu erkennen ist.

Zuletzt wurde auf Basis einer weiteren Erkenntnis der Vorverarbeitung eine dritte Landkarten-basierte Visualisierung erstellt. Bei der Analyse der Spalte „entry_details“ ist aufgefallen, dass auch Flüge mit Zwischenstopps vorliegen. Um diese auch, ähnlich wie in der zweiten Visualisierung geographisch mit den Routen anzuzeigen, wurde eine weitere Spalte erstellt, welche die Route darstellt. Hierbei ist diese Visualisierung auch nach Flug ID filterbar, um leichter einzusehen, wie die Routen der jeweiligen Flüge aussehen.

In Abbildung 3.3 wird die Visualisierung aller Flüge mit Zwischenstopps dargestellt. In dieser Karte ist zu sehen, dass es generell weniger Flughäfen gibt, die bei Flügen mit Zwischenstopps genutzt werden. Auch hier sind nationale und internationale Flüge vorhanden, während in Nordamerika jedoch kein Flughafen genutzt wird.

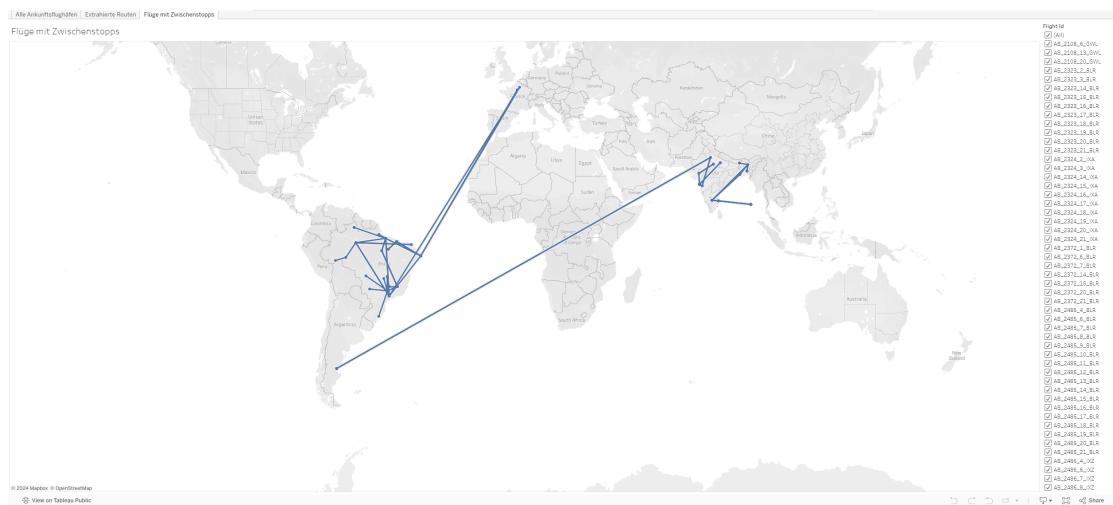


Abbildung 3.3: Landkarte der extrahierten Zwischenstopp

3.2 Weiterführende Visualisierungen

Im Pflichten- und Lastenheft wurde als ergänzende Anforderung das Erstellen von weiterführenden Visualisierungen definiert. Nachdem alle geographischen Insights abgedeckt wurden, war das Ziel dieser Anforderung das Darstellen von weiteren Daten. Im Laufe der Bearbeitung des Projektes ist aufgefallen, dass diese Anforderung sich mit einer weiteren Anforderung überschneidet. Die Ergebnisse aus der Umsetzung der Anforderung zur Prozessabbildung werden in Form von Dashboards abgebildet, welche relevante Daten und Prozesse darstellen. Hierbei beinhalten die Dashboards neben den Prozessen auch bereits Visualisierungen, welche das Ziel dieser Anforderung abdecken. Ein noch nicht abgedeckter Bereich sind dabei die geographischen Daten, welche in Anforderung 1 bereits erfüllt wurden. Dadurch werden, ergänzend zu den Landkarten, in dieser Anforderung weitere Visualisierungen erstellt, welche die Daten der Flughäfen genauer anzeigen.

Als Fokus werden hierbei die Flüge mit einem Zwischenstopp definiert. Das Ziel ist es hierbei einen genaueren Überblick über den Zwischenstopp zu bekommen und genauere Daten der Flughäfen zu sammeln. Insgesamt wurden hierfür drei Balkendiagramme und eine passende Landkarte erstellt, welche die Insights zusätzlich visualisiert. Obwohl

in diesem Schritt der Fokus auf anderen Visualisierungsarten gelegt wird, stellt die Landkarte hier eine sinnvolle Ergänzung dar, um die Erkenntnisse mit den anderen Landkarten zu vergleichen. Alle erstellten Visualisierungen wurden hierbei in einem Dashboard dargestellt, welches in Abbildung 3.4 vorliegt.

In der oberen Hälfte sind die Flughäfen, welche für einen Zwischenstopp verwendet wurden, abgebildet. Zusätzlich zeigt die Größe der Punkte die Anzahl von Flügen an, welche an dem Flughafen einen Zwischenstopp hatten. Generell ist zu sehen, dass die Zwischenstopps hauptsächlich in Südamerika und Asien stattgefunden haben. In der unteren Hälfte beschreiben drei Balkendiagramme die Daten genauer. Zum einen wird in der ersten Graphik ersichtlich, dass neben Indien und Brasilien auch ein Zwischenstopp in Frankreich stattgefunden hat. Zudem fällt auch auf, dass die meisten Zwischenstopps eindeutig in Indien vorkamen. In dem zweiten Balkendiagramm ist die Verteilung der Flüge nach Flughafen zu sehen. Hierbei ist der Flughafen „MAA“ mit knapp einem Viertel der 428 Flüge an erster Stelle. In der dritten Graphik ist zuletzt noch die Aufteilung nach Airline zu erkennen. In diesem Diagramm ist ersichtlich, dass die Airline MN keine Flüge mit Zwischenstopps vorzeigt und dass die Airline AB mit 329 Flügen klar die meisten Zwischenstopps besitzt.



Abbildung 3.4: Dashboard mit weiteren Visualisierungen

Dieses Dashboard zeigt somit mithilfe von weiteren Visualisierungen einen besseren Einblick über die Flüge mit Zwischenstopps. Die Ergebnisse aus diesem Kapitel können somit zusammen mit den Ergebnissen von den Dashboards aus der Anforderung in Abschnitt 5.2 die Anforderung 2 erfüllen.

4 Gewichtswertanalyse

Die Gewichtswertanalyse umfasst sämtliche Fragestellungen für deren Beantwortung die in den Daten vorhandenen Gewichtswerte herangezogen werden müssen.

Im Pflichen- und Lastenheft wurden diesbezüglich drei Anforderungen gestellt:

1. **Grundlagenanalyse** (Anforderung 5): Ziel ist es, sämtliche relevanten Kennzahlen zu identifizieren, deren technische Bedeutung zu erörtern und sie zueinander in Relation zu stellen.
2. **Faktorenanalyse** (Anforderung 6): Ziel ist es, zu ermitteln, welche Faktoren statistisch gesehen einen Einfluss auf die Höhe der Differenz zwischen dem geschätzten und tatsächlichen Abfluggewichts (ETOW & ATOW) haben.
3. **Schätzwertoptimierung** (Anforderung 7): Ziel ist es, ein Machine-Learning-Modell zu trainieren, welches basierend auf initialen Informationen (z. B. Anzahl der Passagiere, Flugzeugtyp, Airline etc.) eine Schätzung des TOW zurückgibt, sodass die Differenz zum ATOW minimiert wird.

Die erste und dritte Anforderung wurden gänzlich erfüllt. Bezuglich der zweiten Anforderung kamen hingegen interne Zweifel auf, ob eine entsprechende Analyse als zwingende Voraussetzung für das Training eines performanten Machine-Learning-Modells angesehen werden muss. Außerdem wurde zu Bedenken gegeben, dass mögliche Erkenntnisse aus dieser Anforderungsbearbeitung wenig Mehrwert für den Auftraggeber bieten. Es wurde sich schlussendlich dazu entschieden, die Anforderung durch folgende zu ersetzen:

Flugzeugtypen-Clustering (Ersetzt Anforderung 6): Ziel ist es, die verschiedenen Flugzeugtypen anhand mehrerer Dimensionen wie deren Gewicht oder deren maximaler Flugdauer miteinander zu vergleichen.

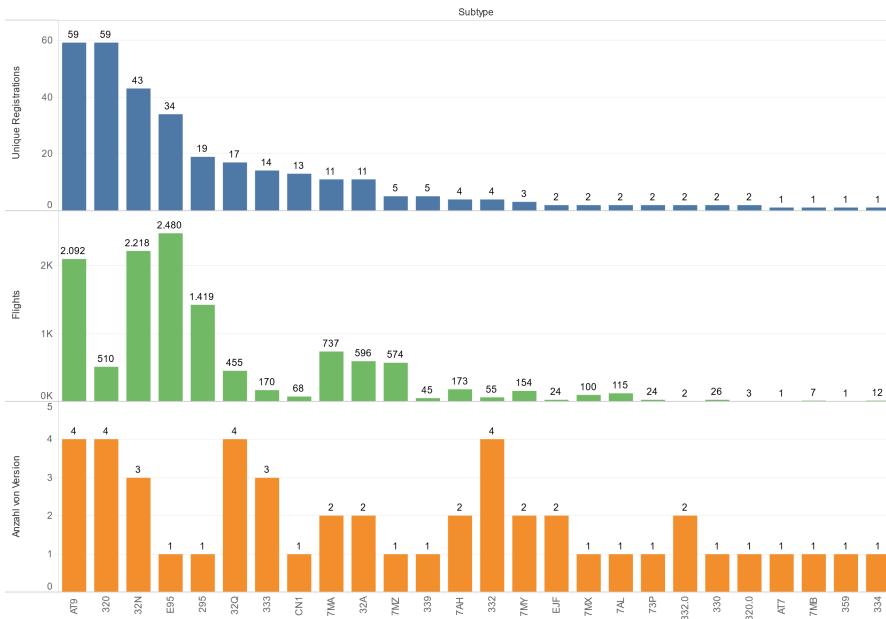


Abbildung 4.1: Größe der Flugzeugflotte (blau) sowie Anzahl durchgeföhrter Flüge (grün) und existierender Versionen (orange) pro Untertyp

4.1 Flugzeugtypen-Clustering

Die drei entscheidenden Features zur Differenzierung verschiedener Flugzeugtypen befinden sich in der Action *ASMMMsgProcessor* und heißen *aircraftSubtype*, *aircraftVersion* und *aircraftRegistration*. Wir gehen davon aus, dass es sich bei der *aircraftVersion*, um bestimmte Varianten eines *aircraftSubtype* handelt und dass die *aircraftRegistration* als einzigartige Identifikationsnummer einzelner Flugzeuge dient. Auffällig ist, dass kein *aircraftType* angegeben wurde. Somit konnte selbst eine umfangreiche Online-Recherche der einzelnen Untertypen und Versionen keine Gewissheit darüber schaffen, mit welcher Art von Flugzeug die einzelnen Flüge letztendlich absolviert wurden.

Während der Analyse galt es verschiedene Herausforderungen zu lösen. Beispielsweise wurde unter der Prämisse agiert, dass jedem Flug genau ein Flugzeug zugeordnet ist, welches über dessen Identifikationsnummer verfolgt werden kann. Tatsächlich ist es aber so, dass sich das zugeordnete Flugzeug vor Beginn des Boarding-Prozesses mehrfach ändern kann. Ebenfalls problematisch war die Tatsache, dass manche Flüge (z. B. solche mit *aircraftVersion* = F0C0Y0) keine Identifikationsnummer besaßen. Diese

und weitere unerwartete Eigenschaften des Datenmodells sorgten für Verzögerungen im Projektfortschritt, konnten aber frühzeitig erkannt und behoben werden.

Über alle drei Fluggesellschaften hinweg gibt es 12.061 Flüge, 263 Flugzeuge und 48 einzigartige Flugzeugmodelle (Untertyp + Version). In Abbildung 4.1 ist zu sehen, wie viele Flugzeuge, Flüge und Versionen es jeweils pro Untertyp gibt.

Die meisten Flugzeuge im Datensatz gehören zum Untertyp AT9, gefolgt von den Typen 320 und 32N. Die meisten Flüge hingegen wurden mit der E95 durchgeführt, gefolgt von 32N und AT9. Auffällig ist, dass der Untertyp 320 trotz einer Flotte von 59 Flugzeugen nur bei 510 Flügen zum Einsatz kam. Die Anzahl der Versionen jedes Untertyps variiert von 1 bis 4 und ist relativ gleichmäßig verteilt.

In Abbildung 4.2 sind die einzelnen Flugzeug-Untertypen in einem Streudiagramm abgebildet. Auf der x-Achse ist das durchschnittliche Zero Fuel Weight (ZFW) der einzelnen Untertypen angegeben, während die y-Achse die durchschnittliche maximale Distanz der Untertypen bemisst. Um diese zu ermitteln, wurde für jeden Flug die Distanz zwischen Abflugs- und Ankunftsflughafen berechnet und das Maximum pro Flugzeugmodell ermittelt. Somit entspricht der Durchschnitt der Maxima pro Untertyp dem Wert auf der y-Achse.

Wie aus der Abbildung hervorgeht, unterscheiden sich die Untertypen erheblich in ihrem Gewicht. Während Untertypen, die mit „33“ und „35“ beginnen, über 100.000 kg wiegen (vermutlich Langstrecken-Passagierflugzeuge wie die Boeing 777), wiegen die verbleibenden Untertypen weniger als 60.000 kg (vermutlich Kurz- bis Mittelstreckenflugzeuge). Obwohl es im Allgemeinen zutrifft, dass schwerere Untertypen längere Strecken fliegen können, gibt es einige Ausnahmen wie die Untertypen 7MX und 7MY, die trotz ihres relativ niedrigen Gewichts von etwa 40.000 kg lange Strecken von über 10.000 km zurücklegen können. Die längsten Nonstop-Flüge der Welt (z. B. von Singapur nach New York) betragen heute etwa 15.000 km. Daher gibt es vermutlich einen Dateneingabefehler oder es handelt sich um ein speziell entwickeltes Flugzeug mit wenig Fracht- bzw. Passagierkapazität. Anhand der Größe der Punkte im Streudiagramm wird die Anzahl der Flüge dargestellt, welche mit dem jeweiligen Untertyp absolviert wurden. So kann man sehen, dass ein Großteil der Flüge mit leichteren Flugzeugen

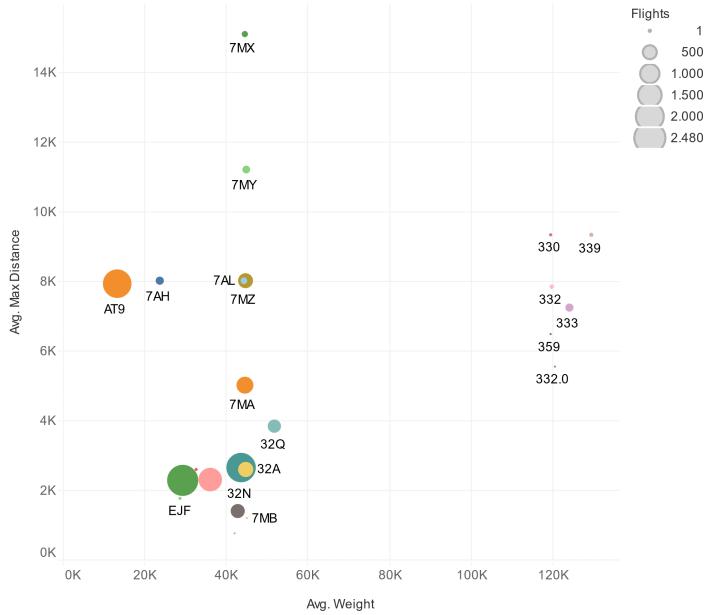


Abbildung 4.2: Reichweite in km und Gewicht in kg sowie Anzahl durchgeföhrter Flüge (Größe) pro Untertyp

(10.000 kg bis 50.000 kg) und über kürzere Strecken (1.500 km bis 8.000 km) durchgeführt wurde.

4.2 Grundlagenanalyse

4.2.1 Identifizierung relevanter Daten

Die Identifizierung der relevanten Daten stellt eine große Herausforderung dar. Schnell ersichtlich ist, dass die *CalculateWeightAndTrimAction* viele relevante Gewichtszahlen enthält. Weniger eindeutig ist, welche anderen Aktionen relevante Informationen enthalten und wie die verschiedenen Aktionen miteinander agieren.

Aus der Prozessanalyse in Kapitel 5 kann entnommen werden, dass die *CalculateWeightAndTrimAction* meist durch eine Änderung der Besatzungszusammensetzung (*UpdateCrewDataAction*), den Check-in eines Passagiers (*CheckinMsgProcessor*) oder der Beladung des Flugzeugs mit Fracht (*UpdateLoadTableAction*) ausgelöst wird. Wir

hatten daher angenommen, dass der Check-in eines Passagiers zu einer Veränderung der Passagiergewichtsdaten führt. Wie der im Anhang befindlichen Abbildung A.1 zu entnehmen ist, ist dies jedoch nicht immer der Fall. Es ist folglich nicht klar ersichtlich, ob und wie die Kennzahlen der einzelnen Aktionen einander beeinflussen.

Doch selbst wenn diese Interaktionsmechanismen bekannt wären, würde eine vollumfängliche Auswertung sämtlicher Aktionen, welche Gewichtsdaten enthalten, aufgrund der Vielzahl an Prozessvariationen viel Zeit und Arbeit in Anspruch nehmen. Daher wurde die Entscheidung getroffen, sich bei der weiteren Analyse auf die beiden Aktionen *CalculateWeightAndTrimAction* und *UpdateLoadTableAction* zu beschränken.

Selbst mit dieser Einschränkung sorgt die Varianz der Prozessabläufe für weitere Herausforderungen. Wie die Abbildungen A.2, A.3 und A.4 zeigen, kann die Anzahl der *CalculateWeightAndTrimAction* pro Flug stark variieren und zwischen 0 und 400 liegen. In der häufigsten Prozessvariation der Fluggesellschaften AB und ZY kommt die Aktion genau einmal vor, während sie bei MN gar nicht auftritt (siehe Kapitel 5). Für diese und viele weitere Flüge ist eine Darstellung der Abweichung zwischen erwartetem und tatsächlichem Abfluggewichts (TOW) also nicht möglich.

Zuletzt ist uns aufgefallen, dass es manche Datenpunkte gibt, welche wir im Rahmen unserer Analyse nicht im Datensatz finden konnten, die jedoch sehr hilfreich wären und den Fluggesellschaften auch sicherlich vorliegen. Ein Beispiel hierfür wäre die Anzahl der Fluggäste, welche sich vorab ein Ticket für den Flug gekauft haben.

4.2.2 Datenverständnis aufbauen

Die in den Aktionen *CalculateWeightAndTrimAction* und *UpdateLoadTableAction* vorliegenden Kennzahlen zu verstehen, hat Zeit in Anspruch genommen. Zunächst haben wir mittels der zur Verfügung gestellten Dokumentation sowie verschiedener Online-Quellen in Erfahrung gebracht, wofür die jeweiligen Akronyme stehen. Anschließend haben wir die Berechnungsvorschriften für die einzelnen Kennzahlen ermittelt. Die hierbei gesammelten Erkenntnisse sind in Abbildung 4.3 zusammengefasst. Sie zeigt,

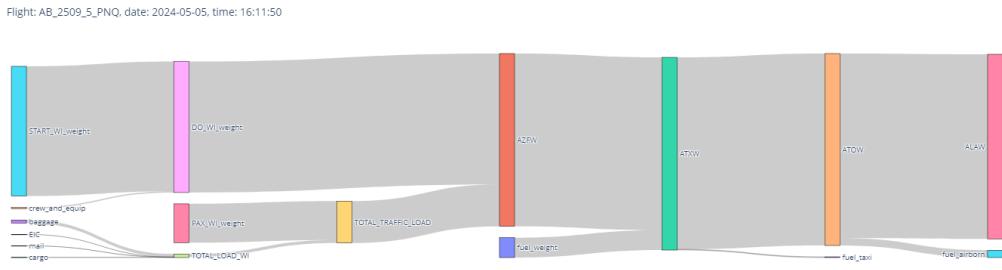


Abbildung 4.3: Visualisierung der Beziehung grundlegender Gewichtsdaten

aus welchen Teilkomponenten sich das ALAW und seine Teilkomponenten zusammensetzen. Hierbei entspricht das Gewicht der jeweiligen Komponente stets der Summe der Gewichte aller direkt vorgelagerten Teilkomponenten.

4.3 Schätzwertoptimierung

Ziel der Schätzwertoptimierung ist es, ein Modell zu trainieren, welches das ATOW möglichst präzise vorhersagen kann. Die erste Herausforderung hierbei war es, herauszufinden, welcher Wert tatsächlich das ATOW eines Fluges darstellt. Die *CalculateWeightAndTrimAction* besitzt ein Feld mit dem Namen ATOW, dieses wird mitunter aber mehrmals pro Flug aktualisiert. Per Definition kann es jedoch nur ein einziges ATOW geben. Wir gehen also davon aus, dass der letzte Wert in diesem Feld das tatsächliche ATOW ist, während es sich bei dem Rest um Schätzungen handelt. Die selbe Logik lässt sich analog auf die Felder AZFW und ALAW anwenden. Es könnte allerdings sein, dass wir mit dieser Annahme falsch liegen, da sonst die Felder ESTIMATED_AZFW und ESTIMATED_TRAFFIC_LOAD redundant wären.

Wenn man die durchschnittliche Abweichung zwischen ESTIMATED_AZFW und AZFW visualisiert, wird zudem ein weiteres Problem ersichtlich. Wie in der ersten Spalte von Abbildung 4.4 zu sehen ist, gibt es extreme Ausreißer im Datensatz. So erhöht sich etwa das AZFW des Fluges AB_1070_1_BOM innerhalb weniger Sekunden von 61.795 kg auf 6.231.070 kg, wobei es sich vermutlich um einen fehlerhaften Datenpunkt handelt.

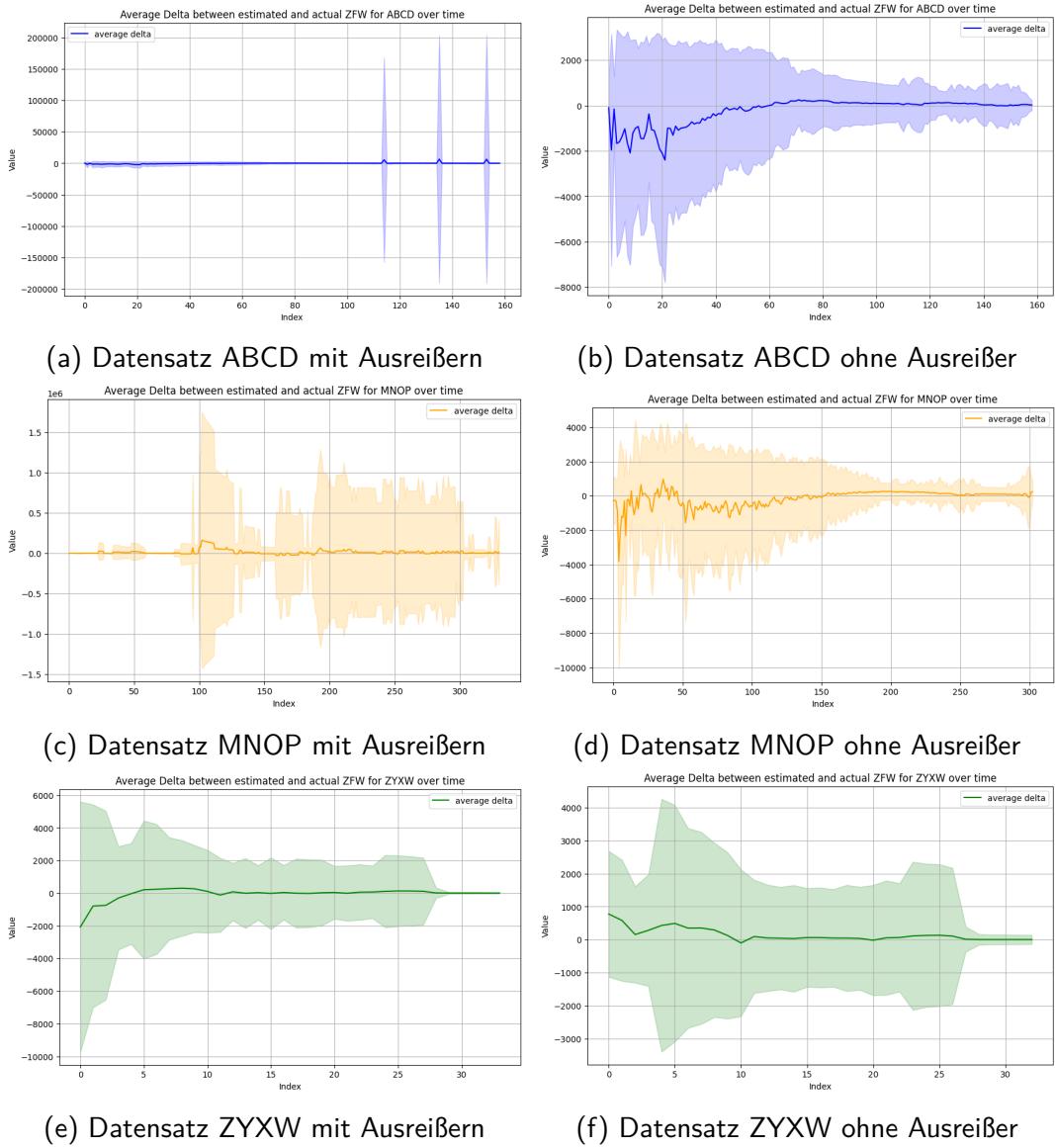


Abbildung 4.4: Mittelwert und Standardabweichung der Differenz aus erwartetem und tatsächlichem ZFW im Zeitverlauf, aggregiert über alle Flüge. Der Graph wurde von rechts nach links generiert. Die statistischen Kennzahlen für den letzten Wert auf der x-Achse wurden aus der letzten *CalculateWeightAndTrimAction* (cawa) des jeweiligen Fluges entnommen. Der vorletzte Wert hingegen aus der vorletzten cawa usw., bis kein Flug mehr über eine weitere cawa verfügt. Demnach basiert die Berechnung der statistischen Kennzahlen auf einer zunehmenden Anzahl von Flügen, je weiter man auf der x-Achse voranschreitet.

In einem ersten Versuch, die Daten zu korrigieren, wurden alle Datensätze entfernt, deren ATOW über 150.000 kg lag. Dies betraf 28.193 Datensätze, wobei 213 der 874 Flüge der Fluggesellschaft MN in Gänze entfernt werden mussten. Auch nach dieser Bereinigung blieben weitere Datenanomalien bestehen. Zum Beispiel gibt es einen kuriosen Flug, der laut Datensatz mit einem Gewicht von 92.741 kg starten und bei der Landung nur noch 7.441.086 kg wiegen soll (flug-id: MN_1086_2_CLE, id: 137785664). In diesem und ähnlichen Flügen mit zweifelhaftem Landegewicht ist stets auch das maximal zulässige Landegewicht (ALLOWED_LAW) in einer ähnlichen Weise verzerrt. Begrenzt man auch die restlichen gewichtsbezogenen Spalten auf ein Maximalgewicht von 150.000 kg, fallen weitere 10.525 Datensätze heraus. Doch nicht nur die Höhe, sondern auch das Verhältnis der Gewichtszahlen untereinander scheint an manchen Stellen fehlerhaft zu sein. So gibt es beispielsweise Flüge, deren ATOW mehr als doppelt so groß ist als das ALLOWED_TOW (flug-id: MN_1432_1DUB, id: 137530320).

In der zweiten Spalte der Abbildung 4.4 ist der Verlauf der Deltas auf den bereinigten Datensätzen zu sehen. Aus ihnen geht hervor, dass die Güte der Gewichtsschätzung im Datensatz zunimmt, je weiter der Beladungsprozess fortgeschritten ist. Die Varianz ist jedoch nach wie vor sehr hoch und es scheint fast sicher, dass der Datensatz mit weiteren Anomalien gespickt ist, welche bislang unerkannt blieben. Diese Mängel in der Güte der Daten werden sich unweigerlich negativ auf die Performance des zu trainierenden Modells auswirken. Da ein Großteil der Mängel bei der Fluggesellschaft MN festgestellt wurde, werden ihre Flüge beim Training nicht berücksichtigt.

Es wurden zwei Modelle trainiert: eine einfache lineare Regression sowie ein neuronales Netz mit zwei hidden Layers der Größe 128 und 64. Als Input dienten die Felder der *CalculateWeightAndTrimAction* sowie einige Informationen aus dem *ASMMsgProcessor* (z. B. *departureAirport*, *aircraftRegistration* etc.) sowie ein berechnetes Feature, welches die Zeit bis zum Abflug in Minuten enthält. Die kategorischen Features wurden per one-hot-encoding transformiert und die numerischen Features mithilfe eines Standardscalers skaliert.

Während die lineare Regression Probleme hatte, die Zusammenhänge in den Daten zu erfassen, konnte das neuronale Netz insgesamt gute Ergebnisse erzielen. Das Training

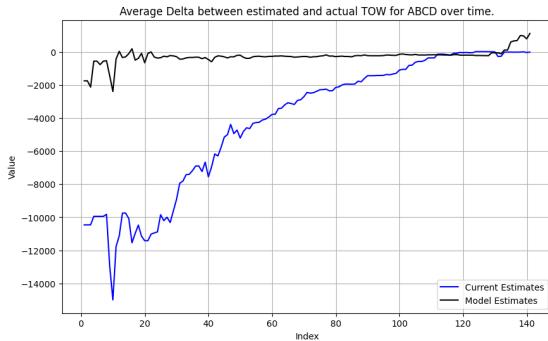


Abbildung 4.5: Durchschnittliche Abweichung des tatsächlichen TOW von den Vorhersagen des neuronalen Netzes sowie der Schätzungen im Datensatz für die Flüge der Fluggesellschaft AB in den Testdaten

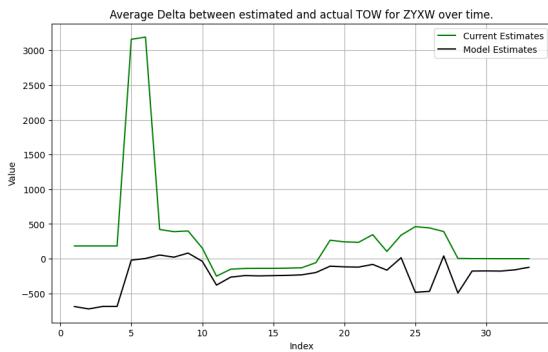


Abbildung 4.6: Durchschnittliche Abweichung des tatsächlichen TOW von den Vorhersagen des neuronalen Netzes sowie der Schätzungen im Datensatz für die Flüge der Fluggesellschaft ZY in den Testdaten

wurde über 100 Epochen und einer Batch Size von 128 durchgeführt. Die Learning Rate betrug zu Beginn 0,1, wurde im Laufe des Trainings aber sukzessive auf einen Wert von 0,0001 reduziert. Zudem wurde das Training in Episode 83 durch Early Stopping abgebrochen.

Die Abbildungen 4.5 und 4.6 zeigen sowohl das Delta zwischen den im Datensatz vorliegenden ETOWs und dem ATOW (blau/grün) als auch das Delta zwischen den Schätzungen des Modells und dem ATOW (schwarz). Wie man sieht, sind die Vorhersagen des Modells stets akkurate als jene, welche im Datensatz zu finden sind. Es gilt jedoch zu beachten, dass es sich bei den Schätzungen um den Durchschnitt über alle Flüge der jeweiligen Airline zum gegebenen Zeitschritt handelt. Da das Modell

einen recht hohen Mean Absolute Error von 2875 aufweist, scheinen die Schätzungen einer hohen Varianz unterworfen zu sein, welche aus der Grafik nicht direkt hervorgeht. Zudem bestehen - wie bereits erwähnt - Zweifel daran, dass die von uns herangezogenen ETOWs tatsächlich den von den Fluggesellschaften verwendeten Schätzwerten entsprechen.

5 Prozessanalyse

Die Analyse des Business-Prozesses in den vorliegenden Daten aus der Flugzeug-Ladeplanung ist ein zentraler Aspekt des Projektauftrages und somit Teil von zwei im Pflichten- und Lastenheft festgelegten Anforderungen. So verlangt Anforderung 3 die Ableitung und Darstellung des allgemeinen Prozesses. Weiter fordert Anforderung 4 die Betrachtung des Automatisierungsgrads der Prozesse sowie eine Analyse der Tätigkeiten, die noch manuell durchgeführt werden.

Auf Grundlage dieser Anforderungen und der verfügbaren Daten wurden die Prozessabläufe identifiziert und analysiert. Hierfür wurden Process-Mining-Verfahren mit der Softwarelösung Celonis angewendet, um die Prozessabläufe der Airlines zu ermitteln und in Form von Prozessdiagrammen darzustellen. Um Rückschlüsse aus den Prozessdaten zu ziehen, wurden weitere KPIs im Detail betrachtet und ausgewertet. Neben der Bereitstellung der Prozessabläufe und der Analyse der historischen Daten wurden für den Kunden mehrere prozessbezogene Dashboards erstellt, die einen verständlichen und übersichtlichen Einblick in die Daten geben und auch in Zukunft mit neueren Daten genutzt werden können.

5.1 Process-Mining

Zur Ableitung der Prozessmodelle wurden die zur Verfügung gestellten Daten zunächst vorverarbeitet. Da es sich bei den Daten um Event-Logs handelt, die eine chronologische Aufzeichnung von Ereignissen darstellen und oft in IT-Systemen verwendet werden, wurde die Kombination aus Airline-ID, Flugnummer, Flugdatum und Abflughafen gewählt, um die Events zu gruppieren. Jede dieser Gruppen, auch bezeichnet als Case, stellt so den Ablauf des Ladeplanungsprozesses für einen bestimmten Flug dar. Der in der Spalte "creation_time" bereitgestellte Zeitstempel gibt den genauen Zeitpunkt eines Events an und wurde der Process-Mining-Software übergeben, um die Reihenfolge und Dauer der Prozessschritte zu ermitteln. Jedes Event hat zudem einen „action_name“,

welcher die durchgeführte Aktivität bezeichnet und im erzeugten Prozessmodell dargestellt wird. Um Komplikationen durch Dopplungen von Zeitstempeln entgegenzuwirken, wurde die ID des Events als sekundäre Sortierspalte eingesetzt. Da Einträge in den Daten doppelt mit unterschiedlichen Kategorien in der Spalte „header_category“ (send, received, saved) abgespeichert werden, muss zur korrekten Darstellung der Prozesse auf eine dieser Kategorien gefiltert werden. Zwar ist Celonis in der Lage, allein mit diesen Spalten ein Process Mining durchzuführen, um jedoch weitere Einblicke und Analysen zu ermöglichen, wurden zudem die weiteren Spalten übergeben.

Mit dem Einsatz von Celonis konnten die Daten analysiert und Prozessmodelle für die verschiedenen Airlines erzeugt werden. Abbildungen dieser Modelle sind im Anhang B.1 zu finden.

Celonis nutzt Process-Mining-Verfahren, um den Prozessablauf zu ermitteln, der in den meisten Fällen auftritt. Es werden somit nicht zwingend alle möglichen Prozessvarianten und Aktivitäten angezeigt, sondern nur diejenigen, die vorherrschend sind. Um einen genaueren Einblick zu bekommen, ist es jedoch möglich, den Detailgrad der Analyse zu erhöhen und einen größeren Anteil an Prozessvarianten zu inkludieren. Hierbei ist jedoch zu beobachten, dass die erzeugten Prozessmodelle schnell sehr groß werden und einzelne Pfade abdecken, die nur in seltenen Fällen auftreten. Dies erschwert eine vernünftige Analyse, und es muss daher ein Kompromiss gefunden werden, der genügend Details beinhaltet, aber das Prozessverständnis nicht behindert.

5.2 Dashboards

Neben der Ableitung und Analyse der Prozessmodelle wurden prozessbezogene KPIs berechnet und die weiteren verfügbaren Daten genutzt. Dadurch konnten neben dem Prozessablauf auch andere Aspekte analysiert werden. Um diese Analysen zu sammeln und insbesondere dem Auftraggeber in verständlicher Weise zu präsentieren, wurden sechs Dashboards erstellt. Diese Dashboards wurden so konfiguriert, dass sie flexibel mit einer anderen Datenquelle wiederverwendbar sind. Sie bieten daher auch den Vorteil, dass der Auftraggeber die Analysen in Zukunft mit neuen, aktuelleren Daten durchführen kann.

Eine Übersicht der Dashboards wird in Abbildung 5.1 dargestellt und eine größere Version jeder dieser Abbildungen ist im Anhang B.2 zu finden.

Alle Dashboards enthalten drei Filter, die es ermöglichen, eine Airline, die richtige „header_category“ und bei Bedarf einen Flughafen auszuwählen. Es ist somit möglich, die Daten flexibel je nach Bedarf zu filtern und die unterschiedlichen Airlines miteinander zu vergleichen.

Das erste Dashboard „**Data Overview**“ erlaubt einen allgemeinen Überblick über die Prozessdaten. Es werden einige zentrale KPIs dargestellt, welche über die Anzahl der ggf. gefilterten Cases und Aktivitäten Auskunft geben. Ein Balkendiagramm vergleicht die Anzahl an Cases der Airlines, eine Tabelle listet alle möglichen Aktivitäten mit ihrem Typ und einer Kurzbeschreibung auf, und eine weitere Tabelle erlaubt die Auswahl und Detailbetrachtung einzelner Cases.

Im „**Process Flow**“ wird der ermittelte Prozessablauf dargestellt. Hier ist ebenfalls eine Auswahl einzelner Cases möglich, und ein Balkendiagramm zeigt den Anteil der Vorkommen der Aktivitäten an. Dieses Dashboard ermöglicht somit einen zentralen Einblick in die Abläufe der Prozesse und erlaubt ebenfalls die Einstellung des gewünschten angezeigten Detailgrads.

Ein weiteres Dashboard unter dem Titel „**Process Time**“ wurde erstellt, um die Durchlaufzeiten der Prozesse zu analysieren. Es wurde hierzu eine Zieldurchlaufzeit von 16 Stunden, also zwei Arbeitstagen, angenommen, diese kann jedoch für zukünftige Analysen flexibel angepasst werden. Das Dashboard gibt neben einigen KPIs zur Erreichung der Zieldurchlaufzeit die Zeiten nach Airline, Datum, Aktivität und Aktivitätstyp an. Mit diesem Dashboard können somit direkt mögliche Probleme in Form von Verzögerungen erkannt und behoben werden.

Ebenfalls von Interesse ist ein Vergleich der Prozessabläufe je nach Airport, da vor allem lokale Gegebenheiten und Unterschiede den konkreten Ablauf des Ladeplanungsprozesses beeinflussen. Das Dashboard „**Airport Drilldown**“ erlaubt genau diese Betrachtung und Analyse. Es werden die Anzahl an Cases, Durchlaufzeiten und prozentuale Erfüllung der Zielzeit, aufgeteilt nach den Flughäfen, dargestellt. Neben einer Tabelle, die diese Werte zur genauen Betrachtung konkret auflistet, bietet zudem eine Darstellung des Prozessdiagramms die Möglichkeit, den konkreten Prozess an genau

einem Flughafen zu betrachten. Diese Informationen ermöglichen es somit, Flughäfen zu identifizieren, welche Abweichungen und Probleme aufweisen.

Ein konkreter Teil der Projektanforderungen war die Analyse des Automatisierungsgrades der Prozesse. Dieser Anforderung wird das Dashboard „**Process Automation**“ gerecht. Da eine Aktivität entweder von einem „human“ oder „service-acco“ durchgeführt wird, können für die Prozesse KPIs in Bezug auf ihren Automatisierungsgrad ermittelt werden. Das Dashboard stellt diese KPIs dar und zeigt, wie viele der Cases und Aktivitäten automatisiert ablaufen. Eine Tabelle gibt zudem mehr detaillierte Einblicke auf Aktivitätsebene.

Zuletzt bietet das Dashboard „**Variant Explorer**“ weitere Einblicke in den Prozessablauf. Während das vorherige „**Process Flow**“-Dashboard zwar ebenfalls den Prozessablauf darstellt und eine Erhöhung des Detailgrads ermöglicht, können hier konkret verschiedene Prozessvarianten ausgewählt und betrachtet werden.



Abbildung 5.1: Übersicht der sechs erstellten Dashboards. Detailansichten sind im Anhang B.2 zu finden.

5.3 Auswertung

Die in Anhang B.1 hinterlegten Abbildungen zeigen den Prozessablauf der verschiedenen Airlines. Zwar unterscheiden sich die Prozessmodelle, bei genauer Betrachtung ist es jedoch möglich, einzelne Elemente einem scheinbaren Hauptprozess zuzuordnen. Dieser besteht aus der Abfolge der folgenden Aktivitäten:

Start - ClearFlightsAction - ASMMsgProcessor - UpdateFlightAction - AssignLLC-CAction - StoreRegistrationAndConfigurationAc - UpdateCrewDataAction - Calculate-WeightAndTrimAction- TransferCargoAction - UpdateFlightAction - ASMMsgProcessor - End

Der Prozessteil zwischen den beiden UpdateFlightAction-Aktivitäten wird hierbei mehrfach, teils auch in abweichender bzw. anderer Reihenfolge, durchlaufen. Die Fluggesellschaften unterscheiden sich durch einige Abweichungen und zusätzliche Aktivitäten. Besonders bei der Erhöhung des Detailgrads wird deutlich, dass dieser Ablauf nur einen Teil der möglichen Prozessvarianten widerspiegelt. Es scheint somit aktuell keinen 100%-tigen Standardprozess, der fest eingehalten wird, zu geben und je nach Flug sind Abweichungen notwendig. Um seine Tätigkeiten zu optimieren, sollte der Auftraggeber somit einen besseren Standardprozess definieren und Schritte unternehmen, um die Anzahl der notwendigen Abweichungen zu reduzieren. Das hat das Potenzial, die Effizienz und Fehleranfälligkeit des Ladeplanungsprozesses zu steigern.

Auf Grundlage dieses durchgeföhrten Process-Minings sowie mit den erstellten Dashboards konnten die Abläufe des aus den Daten hervorgehenden Ist-Prozesses weiter analysiert und eine Auswertung mit mehreren bedeutenden Erkenntnissen erstellt werden.

- Es ist zu erkennen, dass die Airline ZYXW mit 13,5k Cases deutlich mehr Flüge durchführt als die anderen beiden mit jeweils weniger als 3k Cases. Dennoch hat diese Airline mit im Durchschnitt 17 Aktivitäten pro Flugplanung einen scheinbar effizienteren Prozess als die anderen Airlines, welche einen Durchschnitt von 132 und 267 Aktivitäten vorweisen. Hier sei jedoch angemerkt, dass dieser Umstand unter Umständen eher aufgrund von Ausreißern auftritt und die Fluggesellschaft ZY scheinbar viele unvollständige und sehr kurze Prozesse hat, während die anderen beiden einige extrem lange Prozesse vorweisen.
- In Bezug auf die angestrebte Durchlaufzeit von 16 Stunden ist zu erkennen, dass diese im Mittel von den Airlines MNOP (13h) und ZYXW (0,07h) eingehalten werden können. Airline ABCD kann dieser Vorgabe mit 21,6h nicht nachkommen. Die ausschlaggebenden Aktivitäten für diese langen Prozesslaufzeiten unterscheiden sich und es ist kein klarer Faktor für Verzögerungen zu erkennen.

Auch bei dieser Betrachtung sei angemerkt, dass die Durchlaufzeiten scheinbar durch Ausreißer beeinflusst werden und Zeitstempel teils einen fehlerhaften Eindruck vermitteln. Somit ist zu vermuten, dass der eigentliche Prozessablauf der Airline ZYXW deutlich länger ist. Der Auftraggeber sollte daher sicherstellen, dass Zeitstempel korrekt für alle Aktivitäten gesetzt werden und Ausreißer wie „vergessene“ offene Aktivitäten behoben werden.

- Es ist zu beobachten, dass sich die Prozessabläufe nicht nur je nach Airline, sondern vor allem auch je nach Flughafen unterscheiden. Dies ist zu erwarten, da vor allem lokale Gegebenheiten und Gewohnheiten den Prozessablauf beeinflussen. So gibt es einige Flughäfen wie VCP, CNF, DUB und REC, die die meisten Flüge bearbeiten und daher den größten Bedarf für einen effizienten Prozess haben. Weiter ist zu erkennen, dass Flughäfen wie MDZ, BVH und MEU die längsten Durchlaufzeiten haben und es hier besonders viel Potenzial gibt, die Prozessabläufe zu optimieren.
- In Bezug auf die Automatisierung der Prozesse ist positiv anzumerken, dass um die 88% der Cases bereits automatisiert bzw. IT-gestützt durchgeführt werden. Es verbleiben einige Aktivitäten, die immer bzw. sehr oft manuell durchgeführt werden. Dies ist in manchen Fällen zwar ggf. tatsächlich notwendig, jedoch sollte untersucht werden, ob diese Aktivitäten wirklich notwendig sind oder ob sie automatisiert werden können.

Die Analyse der vorliegenden Prozesse konnte tiefe Einblicke in die Betriebsabläufe der Fluggesellschaften geben. Es ist zu erkennen, dass es einen Bedarf für Standardisierung und Fehlerkorrekturen gibt, um die Effizienz der Ladeplanung zu erhöhen. Der Auftraggeber ist mit dieser Analyse und den bereitgestellten Dashboards nun in der Lage, diese Optimierungen anzugehen und Fortschritte oder Probleme benutzerfreundlich und verständlich zu beobachten.

6 Schlussbetrachtung

6.1 Lessons Learned und Reflexion

Neben konkreten Artefakten, die im Rahmen dieses Projekts erstellt wurden, gab es verschiedene Learnings, die reflektiert werden können.

Die Anforderungen der Visualisierungen konnten innerhalb des Projektes erfolgreich erfüllt werden. Bei der Bearbeitung wurde ein Wechsel der verwendeten Tools durchgeführt. Die zu vorherige Entscheidung, Power BI zu verwenden, wurde basierend auf Erfahrungen getroffen, welche im Bereich der Datenvisualisierung gesammelt wurden. Erst nachdem die Ziel-Visualisierungen definiert wurden, ist aufgefallen, dass das Anzeigen der Routen zwischen 2 Punkten in einer geographischen Landkarte in Power BI mit den vorliegenden Daten nicht vollständig möglich ist. Durch die Umstrukturierung kam es zu Doppelarbeit, welche in der Zukunft mithilfe einer Analyse zu Beginn verhindert werden kann.

Die Arbeit mit der Prozessanalyse-Software Celonis hat aufgezeigt, welche umfangreichen Möglichkeiten eine derartige Software bietet. Gleichzeitig wurde die Arbeit aufgrund einer minimalen Dokumentation und der zahlreichen Funktionen, welche teils für eher größere Projekte ausgelegt sind, deutlich erschwert. Ein besonderer Aspekt, der im Laufe der Arbeit immer wieder in den Vordergrund rückte, war zudem die Datenqualität. Während aus den Eventlogs zwar viele Erkenntnisse gewonnen wurden, ist bei einigen Aspekten anzumerken, dass diese durch eine mindere Datenqualität ggf. verfälscht wurden. So deuten viele doppelte Zeitstempel und hohe Variationen in den Durchlaufzeiten darauf hin, dass möglicherweise im Prozessverlauf Daten fehlerhaft abgespeichert worden sein könnten.

Diese Vermutung konnte im Rahmen der Gewichtswertanalyse mehrfach bestätigt werden. Rückblickend haben wir den Fehler gemacht, stets unter der Prämisse eines fehlerfreien Datensatzes zu agieren. So wurde beispielsweise viel Zeit darauf verwendet, die Hyperparameter des neuronalen Netzes anzupassen, um dessen Unzulänglichkeiten

zu beheben. Erst als hier alle Optionen abgeschöpft waren, wurde dessen Datenbasis hinterfragt, wodurch die teils absurd hohen sowie inkongruenten Gewichtswerte entdeckt wurden. Gerade vor dem Hintergrund, dass ein Modell immer nur so gut wie die ihm zur Verfügung stehenden Daten ist, wäre es sinnvoller gewesen, direkt zu Beginn die Datenintegrität zu überprüfen und erst dann mit der Modellierung zu beginnen.

Ein weiteres Learning ist es, dass man bei Projekten mit solch einer großen Datenbasis am Anfang viel Zeit darauf verwenden sollte, einen Überblick über die Daten zu erlangen, um so einen klaren Scope definieren zu können und so den sehr offen gefassten Projektauftrag einzuschränken. Wir haben uns zwar zu Beginn einen umfänglich Überblick verschafft, es jedoch versäumt, unsere Erkenntnisse vernünftig zu Dokumentieren. Es wäre beispielsweise sehr hilfreich gewesen, ein Diagramm zu haben, aus welchem genau hervorgeht, welche Attribute in welcher Aktion zu finden sind und was diese bedeuten. So hätte man eine initialie Auswahl relevanter Aktionen treffen können und hätte sich die Programmierung der Informationsextraktion für irrelevante Aktionen gespart. Zudem hätte dies die Suche der Teammitgliedern nach für sie relevanten Informationen beschleunigt.

6.2 Erfüllte und nicht erfüllte Kriterien

Zu Beginn des Projektes wurden zwischen Auftraggeber und dem Projektteam verschiedene Anforderungen vereinbart und in Form eines Pflichten- und Lastenhefts festgehalten. Mit dem Abschluss des Projektes kann betrachtet werden, welche dieser Anforderungen erfolgreich umgesetzt bzw. nicht vollständig erfüllt werden konnten.

- Anforderung 1: Als Ergebnis der Datenvorverarbeitung konnten die Flughäfen mithilfe von geographischen Visualisierungen dargestellt werden. Hierbei wurden An- und Abflughäfen identifiziert und visualisiert, welche zudem mit den Zwischenstopp-Flügen bereichert wurden. Für die geographischen Visualisierungen wurde eine Tableau Public-Seite erstellt, sodass diese online einsehbar sind.
- Anforderung 2: Für die weiterführenden Visualisierungen wurden bereits erstellte Darstellungen in den Dashboards aus Anforderung 4 verwendet. Zusätzlich

wurde ergänzend ein Dashboard erstellt, welches einen Fokus auf die Flüge mit Zwischenstopps legt und diese genauer betrachtet.

- Anforderung 3: Aus den bereitgestellten Event-Logs wurden mithilfe der Software Celonis Prozessabläufe für die verschiedenen Airlines ermittelt. Zudem konnten auf Grundlage dieser Daten Auswertungen über verschiedene Auffälligkeiten erstellt werden, wodurch die Anforderung der „allgemeinen Prozessabbildung“ erfüllt wurde.
- Anforderung 4: Um eine Analyse der manuellen Tätigkeiten durchzuführen, wurde ein Dashboard mit Einblicken in die Automatisierung der Prozesse erstellt und diese Anforderung damit erfüllt.
- Anforderung 5: Für die Grundlagenanalyse der Gewichtskennzahlen wurden alle zur Verfügung stehenden Daten evaluiert und die Bedeutung sowie Berechnungsvorschriften einiger ausgewählter Kennzahlen ermittelt.
- Anforderung 6: Die angedachte Faktorenanalyse wurde aus den in Kapitel 4 genannten Gründen zugunsten eines Flugzeugtypen-Clusterings verworfen.
- Anforderung 7: Im Rahmen der Schätzwertoptimierung wurden zwei Machine-Learning-Modelle trainiert, wovon eines die Güte der in den Daten vorhandenen Schätzungen übertrifft.

Neben den zu Beginn festgehaltenen Anforderungen wurden im Projektverlauf einige weitere Aspekte umgesetzt, um einen Mehrwert für den Auftraggeber zu bieten. Es wurde im Rahmen der Prozessanalyse nicht nur der Prozessablauf ermittelt, sondern auf Grundlage der Prozessdaten weitere Dashboards erstellt, welche Einblicke in verschiedene prozessbezogene KPIs geben und zudem mit neuen Daten wiederverwendbar sind.

6.3 Ausblick

Das durchgeführte Projekt erfüllt eine Vielzahl an Anforderungen und Analysen auf den bereitgestellten Daten. Es sind weiter noch andere Aspekte denkbar, die neben

Schlussbetrachtung

und auf Grundlage der bisherigen Erkenntnisse umgesetzt werden könnten.

Bei der Betrachtung der geographischen Visualisierungen wäre als Erweiterung eine Kombination mit den analysierten Gewichtsdaten denkbar. Hierbei könnte dargestellt werden, an welchen Flughäfen wie viele Gewichte be- und entladen werden. Zudem könnte diese Darstellung in Verbindung mit den Routen weitere Insights bieten.

Im Rahmen der Prozessanalyse wurde eine ausführliche Betrachtung der Prozesse durchgeführt. Hieraus konnten mehrere Erkenntnisse gewonnen werden; eine Implementierung dieser war jedoch nicht Teil des Projektes. Es ist somit denkbar, diese Erkenntnisse nun in der Realität anzuwenden und weitere Rückschlüsse über das Optimierungspotenzial zu ziehen. Weiter kann von Managementseite ein „korrekter“ Zielprozess definiert werden. Dieser kann dann genutzt werden, um automatisiert in Celonis Abweichungen vom aktuellen Prozess zu identifizieren und konkrete Optimierungen vorzunehmen.

Im Kontext der Gewichtswertanalyse scheint eine Weiterentwicklung der bestehenden Modelle zielführend. Dies könnte etwa über die Erhöhung der Trainingsdauer oder eine Steigerung der Modellkomplexität geschehen. Zunächst sollte allerdings eine erneute Prüfung des Datensatzes auf verbleibende Datenanomalien durchgeführt werden. Des Weiteren wären eine Ausweitung der Input-Features (etwa durch die Einbeziehung weiterer Aktionstypen) oder die Verwendung alternativer Methoden (z.B. Random Forest) denkbar, um die Qualität der Vorhersagen weiter zu verbessern.

A Anhang Gewichtswertanalyse

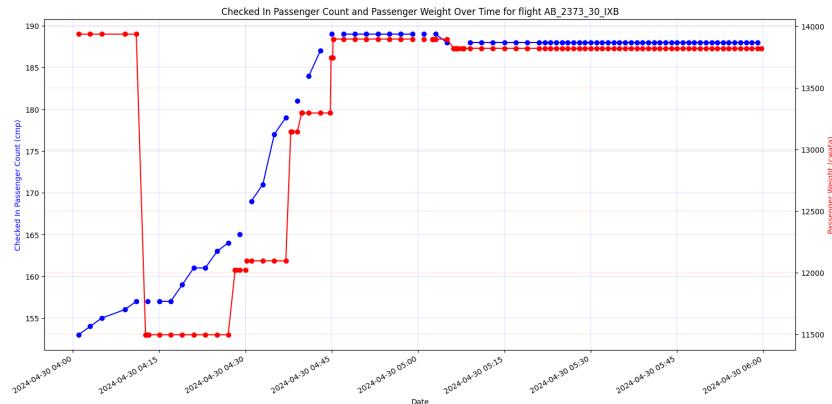


Abbildung A.1: Anzahl der eingekenneten Passagiere (blau, entnommen aus CheckInMsgProcessor) und geschätztes Passagiergewicht (rot, entnommen aus CalculateWeightAndTrimAction)

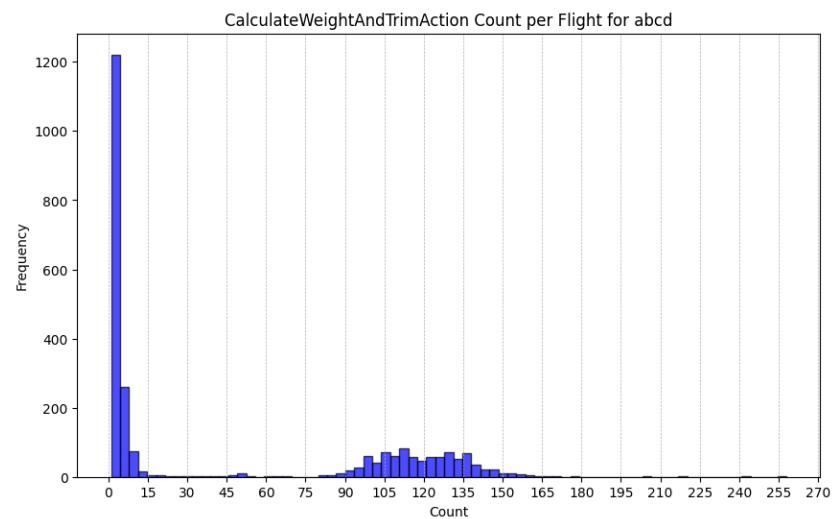


Abbildung A.2: Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft AB. In einem Großteil der Flüge kommt die Aktion weniger als 15 mal vor. Andererseits gibt es auch einige Flüge, welche die Aktion durchschnittlich 120 mal verzeichneten.

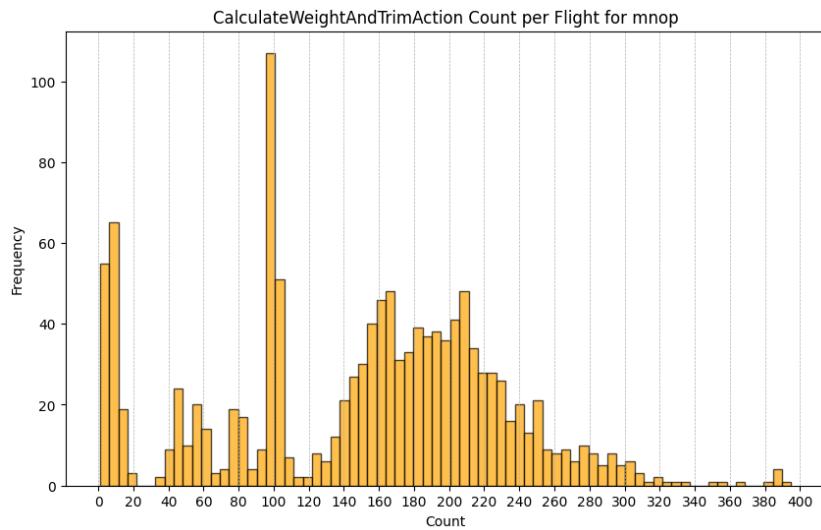


Abbildung A.3: Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft MN. Ansatzweise normalverteilt um 190 mit Ausreißern bei 100 und im Intervall von 0 bis 10.

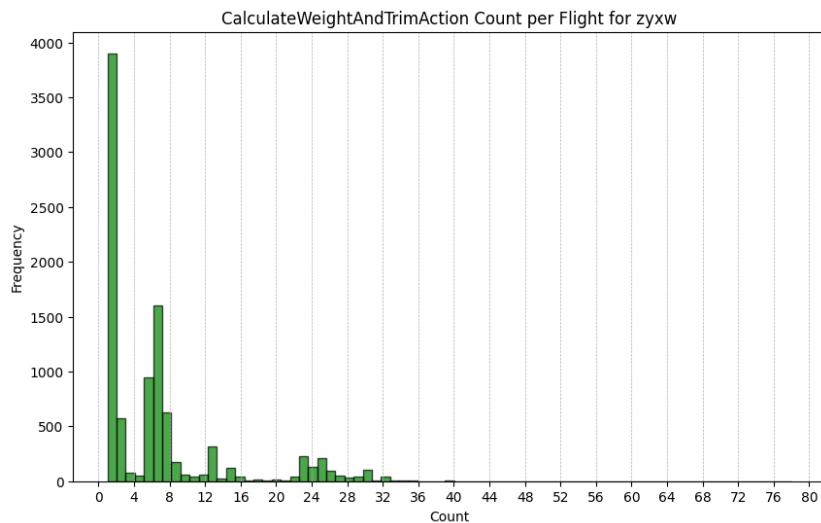


Abbildung A.4: Häufigkeit der CalculateWeightAndTrimAction pro Flug für die Fluggesellschaft ZY. Insgesamt kommt die Aktion hier am seltensten vor.

B Anhang Prozessanalyse

B.1 Prozessablauf

B.1.1 Airline ABCD

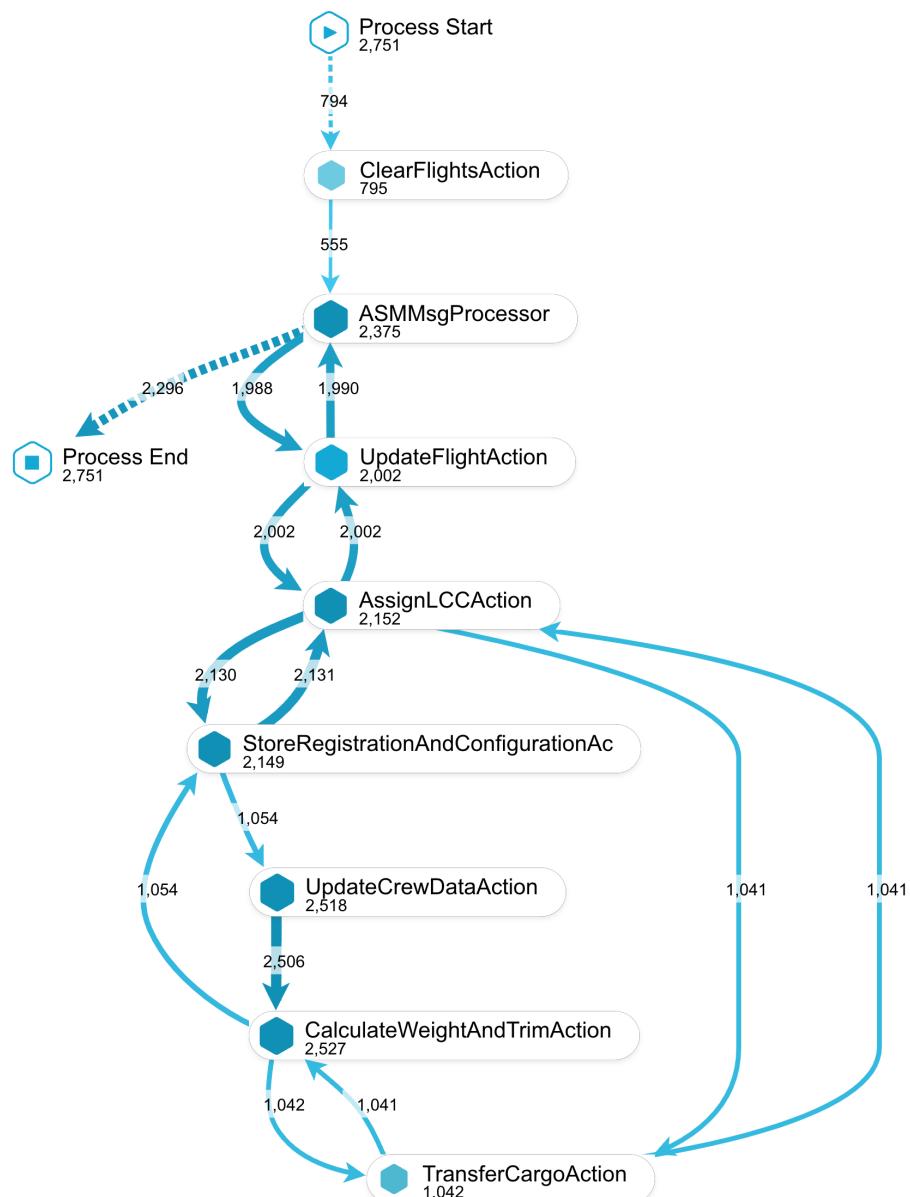


Abbildung B.1: Der häufigste Prozessablauf der Fluggesellschaft ABCD (42,4% der Aktivitäten)

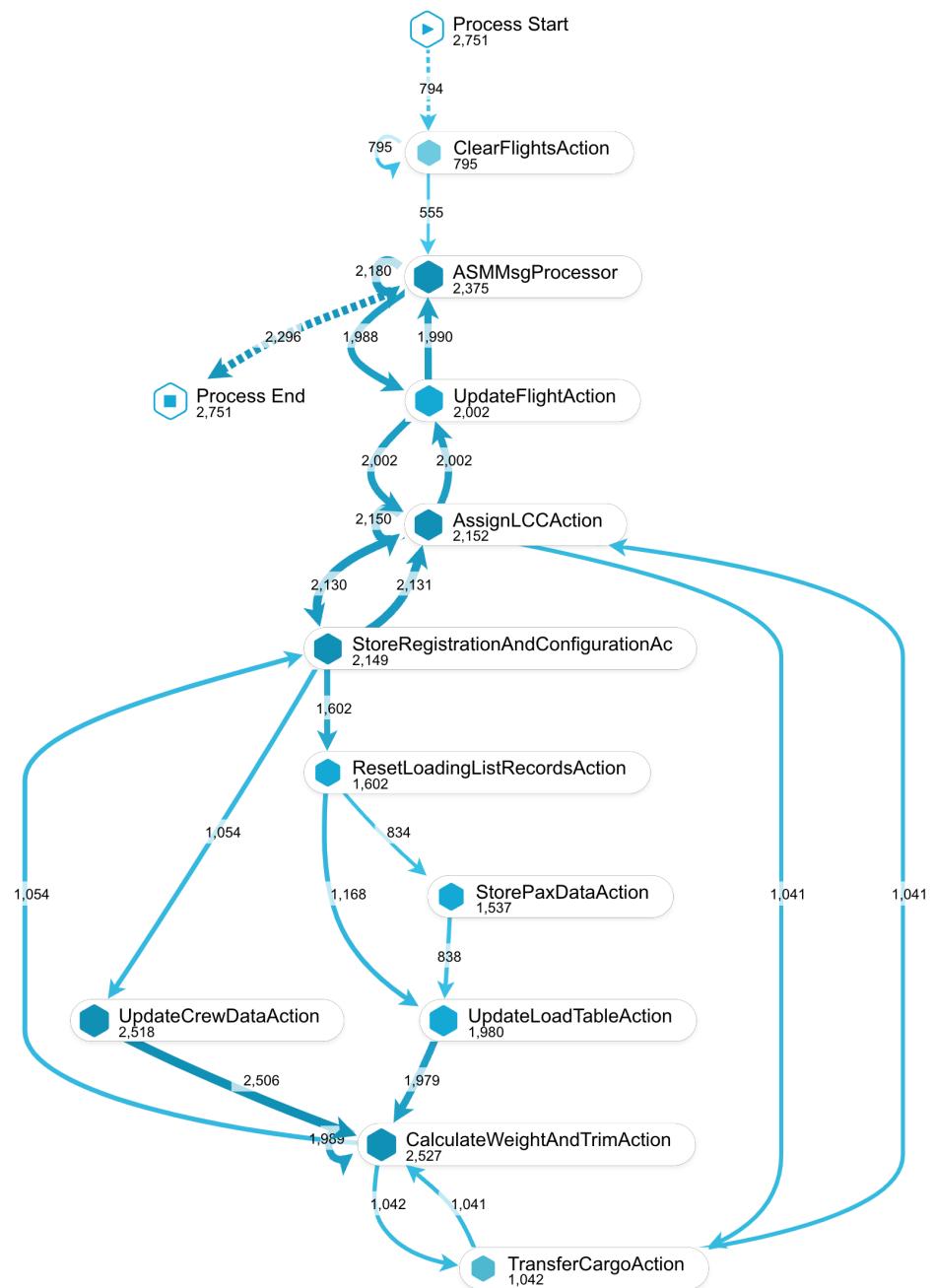


Abbildung B.2: Prozessablauf der Fluggesellschaft ABCD mit erhöhten Detailgrad (67,4% der Aktivitäten)

B.1.2 Airline MNOP

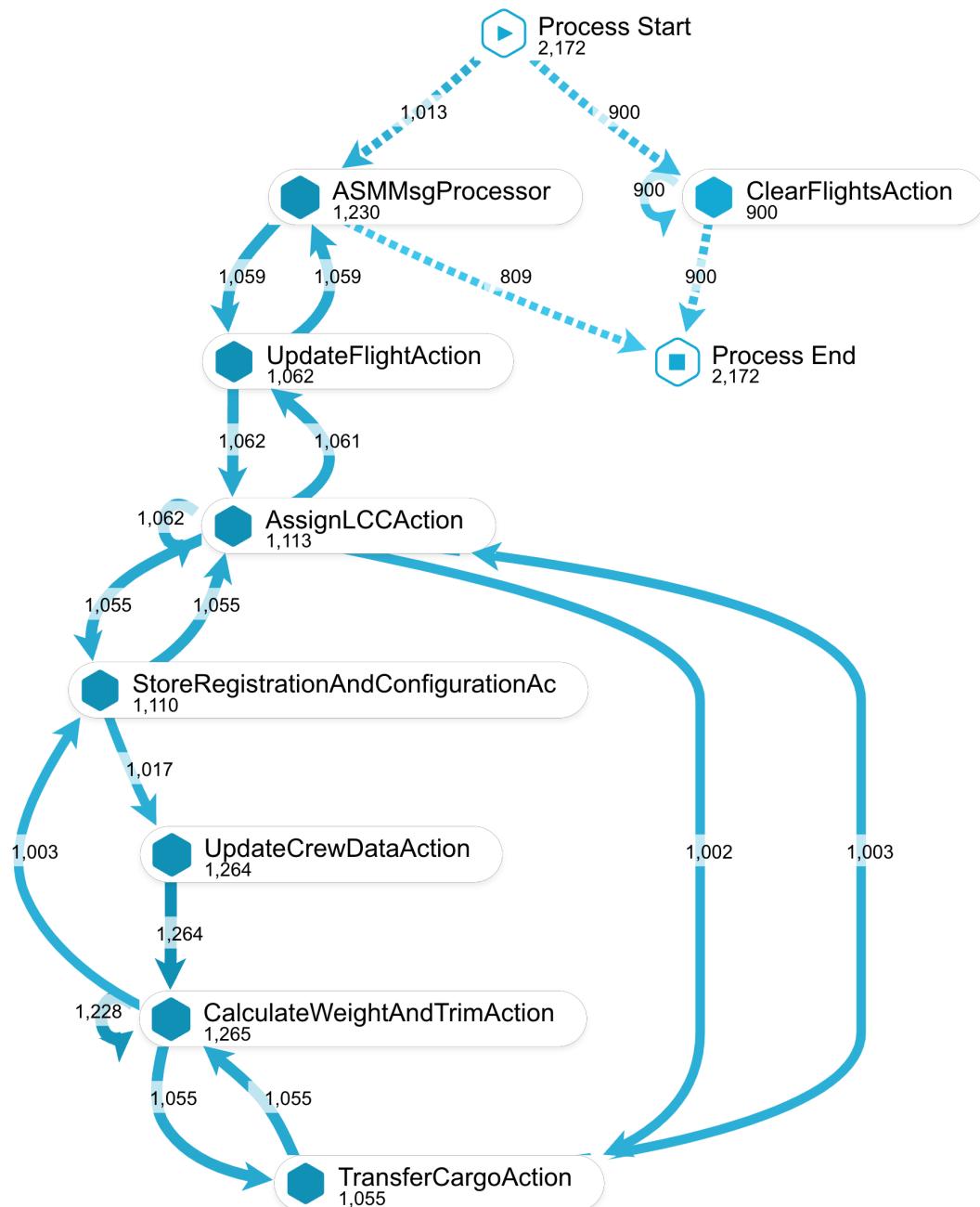


Abbildung B.3: Der häufigste Prozessablauf der Fluggesellschaft MNOP (39,9% der Aktivitäten)

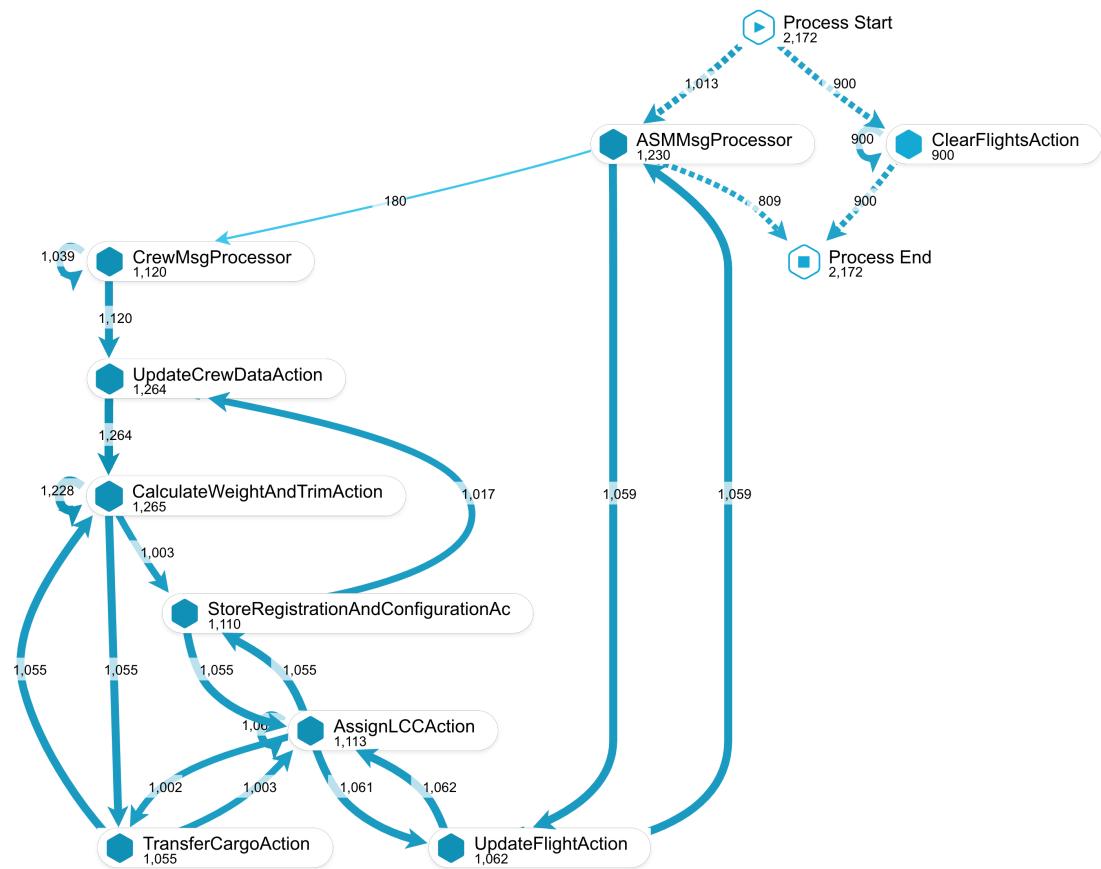


Abbildung B.4: Prozessablauf der Fluggesellschaft MNOP mit erhöhten Detailgrad (45,4% der Aktivitäten)

B.1.3 Airline ZYXW

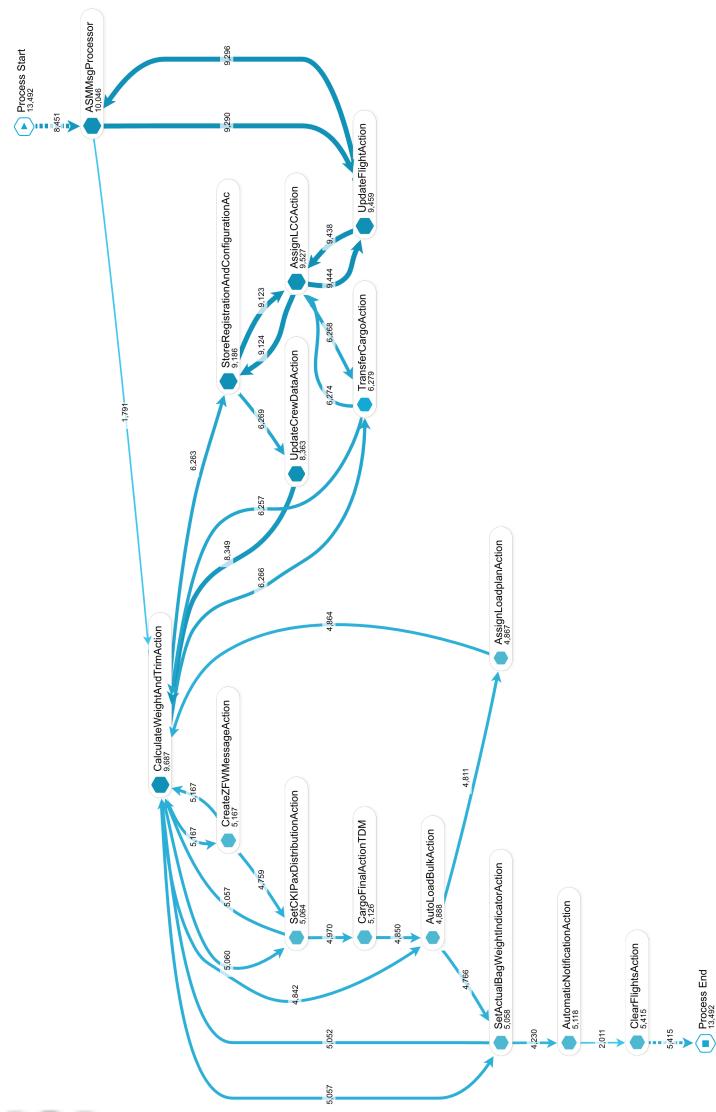


Abbildung B.5: Der häufigste Prozessablauf der Fluggesellschaft ZYXW (85,5% der Aktivitäten)

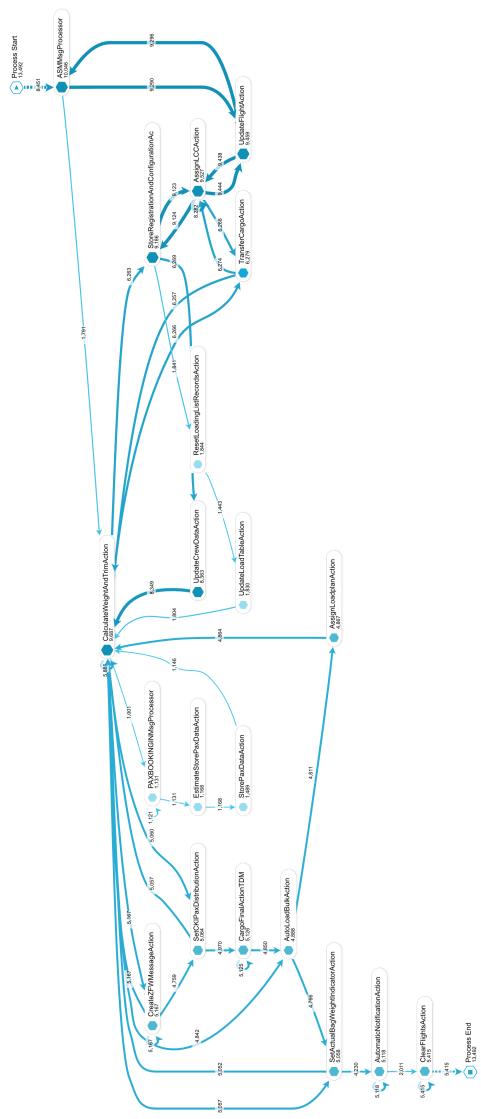


Abbildung B.6: Prozessablauf der Fluggesellschaft ZYXW mit erhöhten Detailgrad (97,1% der Aktivitäten)

B.2 Dashboards

Anhang Prozessanalyse

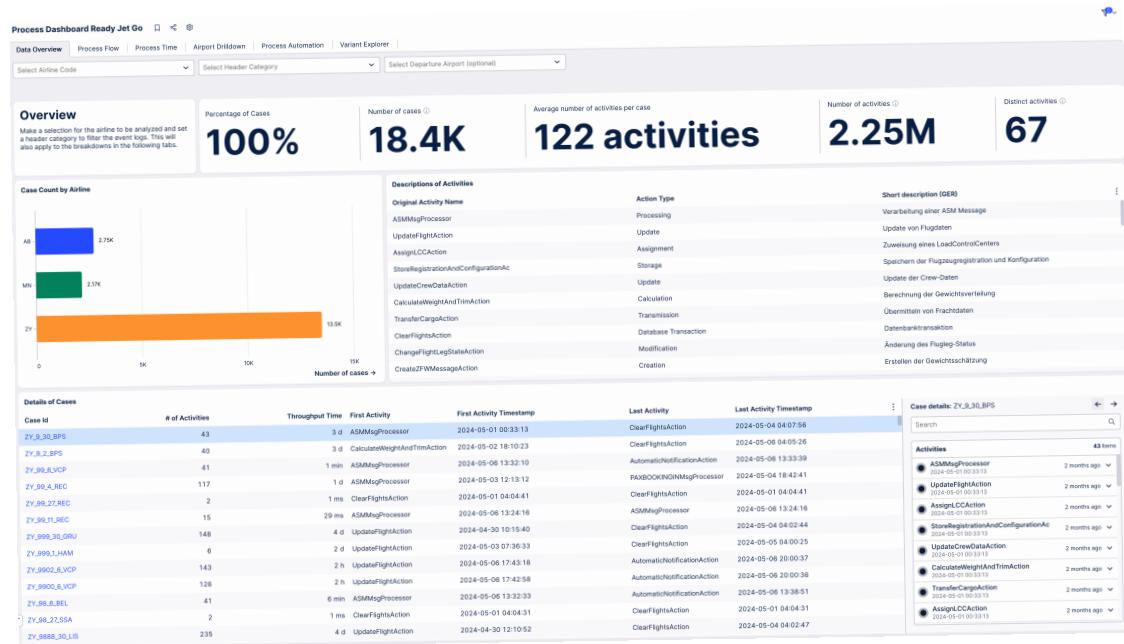


Abbildung B.7: Dashboard „Data Overview“ mit einem allgemein Überblick der relevanten KPIs.

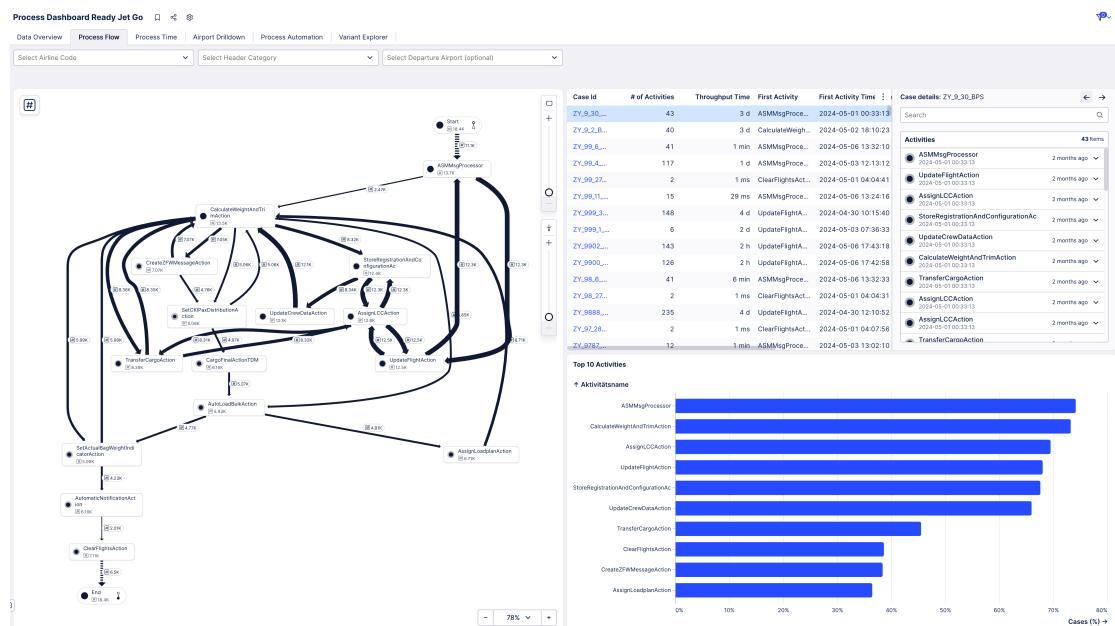


Abbildung B.8: Dashboard „Process Flow“ mit einer Darstellung des Prozessdiagramms.

Anhang Prozessanalyse

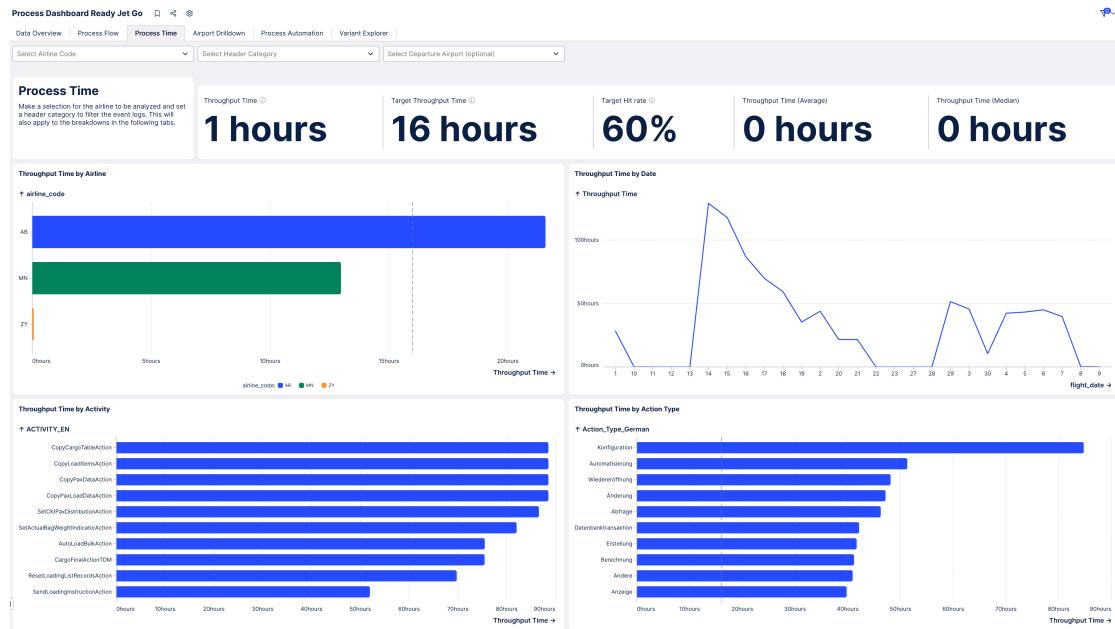


Abbildung B.9: Dashboard „Process Time“ mit Einblicken in die Durchlaufzeiten.

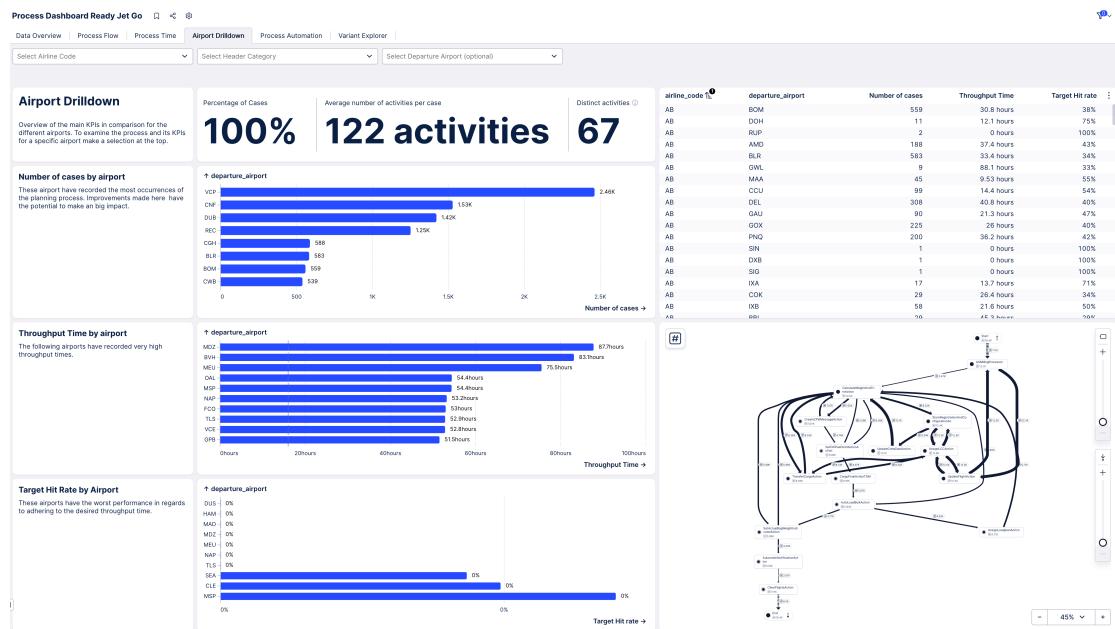


Abbildung B.10: Dashboard „Airport Drilldown“ mit Vergleich der Flughäfen.

Anhang Prozessanalyse

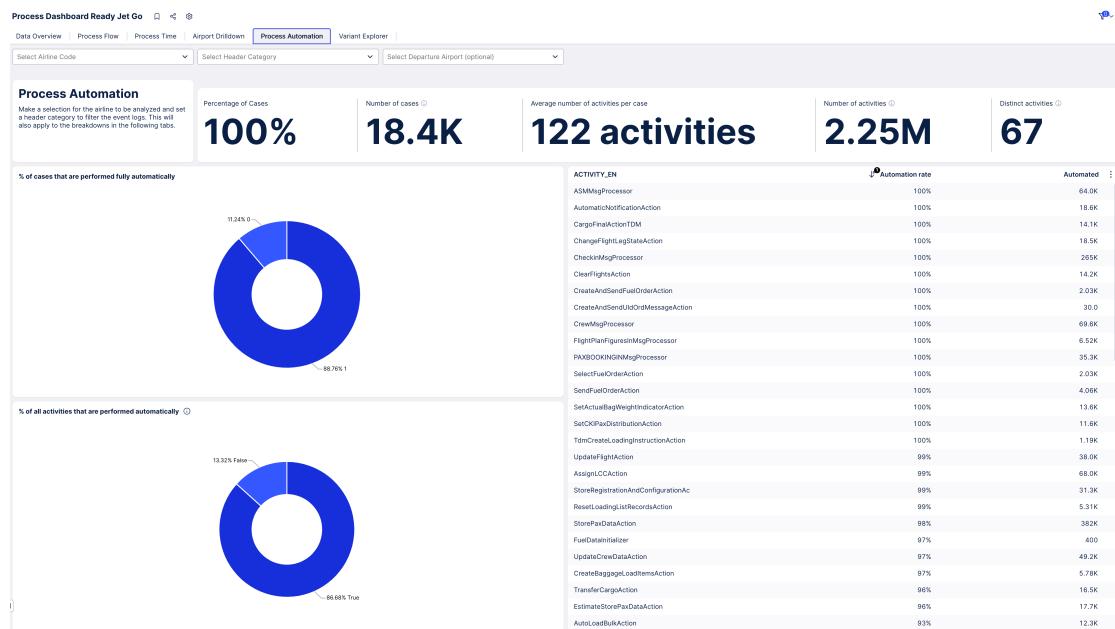


Abbildung B.11: Dashboard „Process Automation“ mit Darstellungen des Automatisierungsgrads.

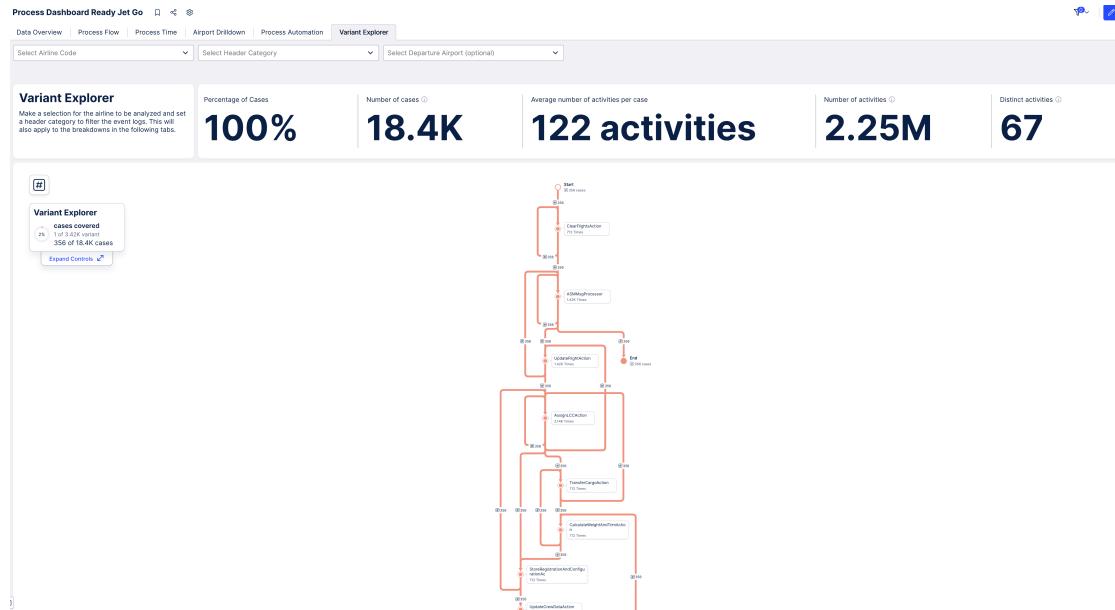


Abbildung B.12: Dashboard „Variant Explorer“ mit Darstellung von Prozess Varianten.

Literaturverzeichnis

- [1] Deutsche Lufthansa AG. *Lufthansa - Flüge, Fluginfo, Check-in, Meilen & Mehr.* Deutsche Lufthansa AG. 2024. URL: <https://www.lufthansa.com> (besucht am 17.07.2024).
- [2] Statista. *Ausschöpfung von Daten in deutschen Unternehmen.* Statista. 2023. URL: <https://de.statista.com/statistik/daten/studie/1387137/umfrage/auschoepfung-daten-in-deutschen-unternehmen/> (besucht am 17.07.2024).