

Below is a comprehensive document listing every engineered feature grouped by its category, along with:

- The exact calculation formula
- The source column(s) from the provided CSV tables

Categories:

1. Ground Time Constraints
2. Traffic & Congestion
3. Operational Complexity
4. Aircraft & Route
5. Passenger Service Needs
6. Process Dependencies

1. Ground Time Constraints

Features that measure turnaround time pressure and buffer:

Feature	Calculation	Source Columns
ground_time_pressure	$(\text{scheduled_ground_time_minutes} - \text{minimum_turn_minutes}) \div \text{minimum_turn_minutes}$	Flight Level: scheduled_ground_time_minutes, minimum_turn_minutes
tight_turnaround	1 if $\text{scheduled_ground_time_minutes} \leq 1.2 \times \text{minimum_turn_minutes}$, else 0	Flight Level: scheduled_ground_time_minutes, minimum_turn_minutes
ground_time_buffer	$\text{scheduled_ground_time_minutes} - \text{minimum_turn_minutes}$	Flight Level: scheduled_ground_time_minutes, minimum_turn_minutes

2. Traffic & Congestion

Features that capture airport and airspace load:

Feature	Calculation	Source Columns
traffic_density	Count of flights sharing same scheduled_departure_date_local & departure_hour	Flight Level: scheduled_departure_date_local, scheduled_departure_datetime_local
airport_congestion	Count of flights sharing same scheduled_departure_station_code & departure_hour	Flight Level: scheduled_departure_station_code, scheduled_departure_datetime_local
congestion_impact	$\text{airport_congestion} \times \text{traffic_density} \times \text{is_peak_hour}$	Derived: airport_congestion, traffic_density, is_peak_hour
is_peak_hour	1 if departure_hour $\in \{6,7,8,16,17,18\}$, else 0	Flight Level: scheduled_departure_datetime_local
rush_hour_intensity	2 if departure_hour $\in \{7,8,17,18\}$; 1 if $\in \{6,9,16,19\}$; else 0	Flight Level: scheduled_departure_datetime_local

3. Operational Complexity

Features that combine multiple stress factors:

Feature	Calculation	Source Columns
stress_factor	$\text{ground_time_pressure} \times \text{load_factor} \times \text{altitude_complexity} \times \text{process_parallelism}$	Derived: ground_time_pressure, load_factor, altitude_complexity, process_parallelism
altitude_complexity	3 if departure_hour \in peak; 2 if \in shoulder; else 1	Flight Level: scheduled_departure_datetime_local
operational_risk	$\text{tight_turnaround} \times \text{congestion_impact} \times (\text{service_complexity} + \text{baggage_complexity})$	Derived: tight_turnaround, congestion_impact, service_complexity, baggage_complexity
prev_flight_risk	previous flight's operational_risk for same fleet_type	Derived via shift on fleet_type_encoded group

4. Aircraft & Route

Features describing aircraft capacity and route complexity:

Feature	Calculation	Source Columns
---------	-------------	----------------

fleet_type_encoded	LabelEncoder transformation of fleet_type	Flight Level: fleet_type
destination_encoded	LabelEncoder transformation of scheduled_arrival_station_code	Flight Level: scheduled_arrival_station_code
origin_encoded	LabelEncoder transformation of scheduled_departure_station_code	Flight Level: scheduled_departure_station_code
is_mainline	1 if carrier == 'Mainline', else 0	Flight Level: carrier
scheduled_flight_duration	(scheduled_arrival_datetime_local – scheduled_departure_datetime_local).minutes	Flight Level: scheduled_arrival_datetime_local, scheduled_departure_datetime_local
total_seats	As-is	Flight Level: total_seats

5. Passenger Service Needs

Features capturing passenger-related complexity:

Feature	Calculation	Source Columns
load_factor	$\text{total_pax} \div \text{total_seats}$	PNR Flight Level: total_pax; Flight Level: total_seats
high_touch_pax_ratio	$(\text{wheelchair_requests} + \text{children} + \text{lap_children} + \text{stroller_users}) \div \text{total_pax}$	PNR Remarks: wheelchair_requests; PNR Flight Level: children, lap_children, is_stroller_user; PNR aggregate: total_pax
basic_economy_ratio	$\text{basic_economy_pax} \div \text{total_pax}$	PNR Flight Level: basic_economy_pax; PNR aggregate: total_pax
ssr_per_pax	$\text{ssr_count} \div \text{total_pax}$	PNR Remarks aggregate: ssr_count; PNR aggregate: total_pax

6. Process Dependencies

Features measuring operational coordination efficiency:

Feature	Calculation	Source Columns
---------	-------------	----------------

process_parallelism	$\min(\text{baggage_complexity} \div (\text{boarding_complexity} + 0.1), 2.0)$	Derived: baggage_complexity, boarding_complexity
boarding_complexity	$(\text{total_pax} \div \text{total_seats}) \times (1 + \text{children} \div 100)$	PNR Flight Level: total_pax, children; Flight Level: total_seats
baggage_complexity	$\text{total_bags} \div \text{scheduled_ground_time_minutes}$	Bag Level aggregate: total_bags; Flight Level: scheduled_ground_time_minutes
bags_per_seat	$\text{total_bags} \div \text{total_seats}$	Bag Level aggregate: total_bags; Flight Level: total_seats
bags_per_minute	$\text{total_bags} \div \text{scheduled_ground_time_minutes}$	Bag Level aggregate: total_bags; Flight Level: scheduled_ground_time_minutes

All features were engineered from the provided CSV tables according to the calculations above, then aggregated and merged into the main flights DataFrame using the common flight keys:

```
company_id + flight_number + scheduled_departure_date_local +
scheduled_departure_station_code + scheduled_arrival_station_code
```

This document serves as a complete reference for all operational feature categories, their precise formulas, and data sources.