

Big Data THack 2013 Amadeus Dataset Description





Index

About Amadeus	3
About the dataset	3
In a nutshell	3
Where is data coming from	3
Data available	
Date (mandatory)	4
Time (mandatory)	4
Origin of the trip (mandatory)	5
Destination of the trip (mandatory)	5
Round Trip Indicator (mandatory)	
Number of segments (mandatory)	5
Segment x Departure (mandatory)	6
Segment x Arrival (mandatory)	
Segment x Date (mandatory)	6
Segment x Carrier (optional)	7
Full example	7
Data domain	7
Your Amadeus Point of Contact for support during the event	8

Document control

Security level	Confidential Restricted				
Company	Amadeus IT Group SA				
Department	R&D-ORI-NOV				
Author	Nicolas HAUVILLER				
Approved by	Mary KEAGUL		Date	DD / MM / YYYY	
Approved by	Landry HOLI		Date	DD / MM / YYYY	
Version	Date	Change	Comment	Ву	
1.0	01/10/2013		Initial Release	N.Hauviller	
1.01	02/10/2013	Minor changes		T.Pearson	



About Amadeus

Amadeus is a leading provider of advanced technology solutions for the global travel industry. Customer groups include travel providers (e.g. airlines, hotels, rail and ferry operators, etc.), travel sellers (travel agencies and websites), and travel buyers (corporations and travel management companies).

The Amadeus group employs around 10,000 people worldwide, across central sites in Madrid (corporate headquarters), Nice (development) and Erding (operations), as well as 71 local Amadeus Commercial Organisations globally.

The group operates a transaction-based business model. For the year ended December 31, 2012 the company reported revenues of €2,910.3million and EBITDA of €1,107.7 million.

Amadeus is listed on the Spanish Stock Exchange under the symbol "AMS.MC" and is a component of the IBEX 35 index.

To find out more about Amadeus please visit http://www.amadeus.com

About the dataset

In a nutshell

The dataset provided by Amadeus in the scope of the THack Big Data 2013 refers to the flight search queries (low fare search) made by travelers from Travel Agencies Websites or Online Travel Agencies worldwide, for a 6 month period in 2011.

Where is data coming from

Data corresponds to the traffic linked to Amadeus Low Fare Search solutions known as the **Amadeus Master Pricer Portfolio** (more information here).

Amadeus Master Pricer Portfolio is the most revolutionary and comprehensive low fare search and merchandising solution for **online travel agencies** and **online sites of travel agencies**.

It allows travel agencies' customers to search for flights based on a wide array of criteria, including price, time of departure/arrival and airline for both domestic and international itineraries.

Data available

Each time a single traveler is searching for a flight (trying to get the lowest fare available) a new entry is recorded similar to this one:

Date	01/04/2011
Time	00:42:37
Origin of the trip	BWI
Destination of the trip	CDG
Round Trip Indicator	1
Number of segments	2
Segment 1 Departure	BWI





Segment 1 Arrival	CDG
Segment 1 Date	14/12/2011
Segment 1 Carrier	ВА
Segment 2 Departure	CDG
Segment 2 Arrival	BWI
Segment 2 Date	05/01/2012
Segment 2 Carrier	ВА
Segment 3 Departure	
Segment 3 Arrival	
Segment 3 Date	
Segment 3 Carrier	
Segment 4 Departure	
Segment 4 Arrival	
Segment 4 Date	
Segment 4 Carrier	
Segment 5 Departure	
Segment 5 Arrival	
Segment 5 Date	
Segment 5 Carrier	
Segment 6 Departure	
Segment 6 Arrival	
Segment 6 Date	
Segment 6 Carrier	

Date (mandatory)

Description

Date when the traveler performed the flight search

Format

YYYY-MM-DD where:

- YYYY is the year of the request
- MM is the month of the request
- DD is the day of the request

Example

'2011-06-04'

Time (mandatory)

Description

Time (UTC - Coordinated Universal Time) when the traveler performed the flight search **Format**

hh:mm:ss where:

- hh corresponds to the hour
- mm corresponds to the minutes
- ss corresponds to the seconds

Example

'00:01:30'



Origin of the trip (mandatory)

Description

¹IATA code corresponding to the city or the airport defined as an origin for the flight search (i.e. from where the traveler will travel).

Correspondence between IATA City codes/City names or Airport Code/Airport Name can be found here: http://www.iata.org/publications/Pages/code-search.aspx

Format

[3 alphanumeric characters]

Example

'NYC'

NYC means that the traveler was searching for a flight departing from New York City (whatever airport).

'JFK'

JFK means that the traveler was searching for a flight departing from JFK Airport (and not any other airport in New York City)

<u>Note</u>: The city corresponds to the flight's departure city in the search request and not to the origin of the request itself (even if in most cases, this city will correspond to the "origin of the request").

For instance, NYC means that the traveler is looking for flights from NYC, but is not necessarily living close to NYC (even this is probably the case).

Destination of the trip (mandatory)

Description

IATA code corresponding to the city or the airport defined as a destination for the flight search (i.e. where the traveler wants to travel).

Format

[3 alphanumeric characters]

Example

'PAR'

PAR means that the traveler was looking for a trip to the city of Paris, France (any airport).

Round Trip Indicator (mandatory)

Description

Determine whether the user was searching for a one way flight (0) or a round trip flight (1)

Format

0 or 1

Example

'1'

1 means that the traveler searched for round trip flights.

Number of segments (mandatory)

Description

This field provides information about the number of segments which are part of the request (see below and example).

Format

1 number (from 1 to 6)

Example

'2'

¹ The International Air Transport Association (IATA) is an international industry trade group of airlines whose role is, among other things, to define all the Airline rules and regulations.



Confidential

For each segment (up to 6)

A segment can be defined as one flight between a departure and an arrival. A trip can be defined as a set of segments between an origin and a destination.

E.g. If a traveler is searching explicitly for a round trip from Miami to Paris via New York City, the trip can involve 4 segments because there is one connection:

Segment 1

Departure 1: Miami

Arrival 1: New York City

[Connection]

Segment 2

Departure 2: New York City

Arrival 2: Paris

Segment 3

Departure 3: Paris

Arrival 3: New York City

[Connection]

Segment 4

Departure 4: New York City

Arrival 4: Miami

Segment x Departure (mandatory)

Description

IATA code corresponding to the city or the airport defined as the departure for the segment.

Note: for segment #1, this field always correspond to the "Origin" field

Format

[3 alphanumeric characters]

Example

'ATL'

Segment x Arrival (mandatory)

Description

IATA code corresponding to the city or the airport defined as the arrival for the segment.

Format

[3 alphanumeric characters]

Example

'NYC'

Segment x Date (mandatory)

Description

Date requested for the flight departure

Format

YYYY-MM-DD where:

- YYYY is the year of the flight departure
- MM is the month of the flight departure





DD is the day of the flight departure

Example

'2011-08-01'

- For Segment #1, this date corresponds to the travel date (i.e. when the traveler wants to travel)
- For the latest segment, this date often corresponds to the return date (i.e. when the traveler wants to go back home)

Segment x Carrier (optional)

Description

IATA code corresponding to the traveler's preferred carrier. If this field is not empty, this means that the traveler explicitly specified that he/she wants to travel with a given airline. Correspondence between IATA Airline codes/Airline names can be found here: http://www.iata.org/publications/Pages/code-search.aspx

Format

[2 alphanumeric characters]

Example

'DL'

DL means that the traveler, in his/her search request specified that he/she wants to fly with Delta Airlines for this segment.

Full example

```
2013-07-07,00:00:01,FRA,VLC,1,4,FRA,MAD,2013-10-15,IB,MAD,VLC,2013-10-15,IB,VLC,MAD,2013-10-22,IB,MAD,FRA,2013-10-22,IB,,,,,,,,,
```

This entry means that on 07 July 2013 at 00:00:01 UTC someone searched for travel solutions (using an online travel agency website or the website associated to a traditional travel agency) to go from FRA (Frankfurt, Germany) to VLC (Valencia, Spain) via MAD (Madrid, Spain) on 15 Oct 2013, with a return one week later on 22 Oct 2013. The traveler explicitly specified that he/she wants to travel with IB (Iberia)

Data domain

The dataset provided by Amadeus in the scope of the THack Big Data will have these characteristics:

- Worldwide origins and destinations for search gueries
- Search gueries from the 1st of April 2011 to the 31st of October 2011
- Search queries are provided in Gzipped CSV files
 - A total of 199 files with an average size of 500Mb (zipped) or 4Gb (unzipped)
 - One file per day having the following format: YYYYMMDD.csv.gz
 e.g. 20110401.csv.gz is a Gzipped CSV file with all the search queries done on the 1st of April 2011
 - The first row of each csv file will contain a copyright statement. The remaining rows will be formatted as previously explained





Your Amadeus Point of Contact for support during the event



Tadhg PEARSON tadhg.pearson@amadeus.com