Crew Operations Analytics and Business Intelligence at American Airlines, Dallas Fort Worth, TX

By

Madhura A. Indurkar

A Project report submitted to the Integrated Manufacturing Systems Engineering Institute of North Carolina State University in Partial fulfillment of the requirements for the Degree of

Masters in Integrated Manufacturing Systems Engineering

Raleigh, North Carolina

2018

Approved by:

Dr. Steven D. Jackson
Advisory Committee Chairman
Director, IMSE Institute

Dr. Thom Hodgson
Distinguished University Professor
Edward P. Fitts Department of Industrial and Systems Engineering

Dr. Michael Spano
Associate Professor
Edward P. Fitts Department of Industrial and Systems Engineering

ABSTRACT

INDURKAR, MADHURA. Crew Operations Analytics and Business Intelligence at American Airlines, Dallas Fort Worth, TX. (Under the Guidance of Dr. Steven D. Jackson, 2018)

This report describes in detail my work on one of the two projects that I undertook during my cooperative experience at American Airlines at its corporate headquarters in Dallas Fort Worth, TX. The scope of the project included extraction of the actual and planned flight trips data to identify the Crew operations metrics contributing to increasing operational costs in 2016 and first business quarter of year 2017. The need for this analysis arose due to excessive Crew operational costs and heavy losses in income due to inaccurate forecasts of operational parameters. The aim of the project was to identify the causes of variance and provide insights regarding the forecasts and prediction accuracy of Crew operations by the Analytics team and suggest improvement areas to management for better decision making.

The project was executed in three phases:

1. Mining of Actual and Planned Flight Trip Data for a period of 18 months from multiple tables in a Teradata database, by writing SQL queries.
2. Data preparation, validation and calculation of operational metrics
3. Variance analysis and visualization of operations metrics

The results of the variance analysis provided actionable insights, enabling analysts and stakeholders to formulate a strategy roadmap targeting areas of improvement and ultimately to improve the overall operational performance of the Airline.

ACKNOWLEDGEMENT

The Internship/Co-op experience at American Airlines, Dallas Fort Worth, TX was a great opportunity for learning and professional development. The successful completion of this project would not have been possible without the help and support of a lot of individuals and organizations.

I would like to extend my sincere gratitude to my defense committee chairman, Dr. Steven Jackson for his continuous guidance and support throughout the completion of this project. I would also like thank my committee members Dr. Thom Hodgson and Dr. Michael Spano and the entire faculty at North Carolina State University and the IMSEI department for their support and encouragement.

I also owe a deep gratitude to Mr. Ameya Kingaokar, Manager for the Crew Strategy and Business Intelligence team who gave me the opportunity to work on this project. I would also like to thank all my fellow analysts Miss Ke Sun, Miss Rachael Van Winkle and Mr. Praneeth Kothagattu at American Airlines for their constant guidance and support.

COMPANY INTRODUCTION

American Airlines Inc. is one of largest domestic service providing airlines in the US. And is headquartered in the Dallas Fort Worth area in Texas. American Airlines currently operates ten hubs across the country and is the world's largest airline with respect to number of destinations served, number of passengers carried, fleet size and revenue generated. In 2013, American Airlines merged with US Airways, another well-established major airline in the US.

The airline's parent company American Airlines Group operates an international and domestic network of nearly 6,700 flights per day to nearly 350 destinations spanning across more than 50 countries, in conjunction with its regional partner American Eagle. American Airlines is a founding member of Oneworld alliance, the third largest airline alliance in the world in collaboration with British Airways, Iberia, Finnair, Cathay Pacific and Japan Airways. [https://www.aa.com/i18n/customer-service/about-us/history-of-american-airlines.jsp]

Figure 1: American Airlines Company logo

WORK EXPERIENCE

I was working in the Crew Strategy and Business Intelligence Team which is a part of the Crew Operations Department in the Corporate Technology sector of American Airlines. Apart from stakeholders in management, the team members had diverse roles of technical analysts, statisticians, programmers and consultants. For the purpose and scope of my projects, I also had to collaborate with employees from the Operations Performance Team.

The projects I worked on gave me the opportunity to work on different technical areas ranging from Strategy Development and Business Intelligence to Data mining and analysis using Statistical techniques and Machine Learning models. This practical, hands-on experience gave a more in-depth understanding of airline industry data and its complexity, the analytical practices applied, and software tools used for analyses. It was very exciting to apply the concepts and methods learned in the classroom in a real-life industry scenario.

I also organized a networking event of fellow interns from the Crew Operations Department with management of various departments like Finance, Operations consulting and Revenue Management. The event helped us interns understand the interdependency of these departments, the differences and similarities in the application of data and the role they play in the overall functioning of a major corporation.

Apart from the knowledge and experience gained, I was also able to travel to a lot of places on weekends during the internship period on account of the flight benefits offered to interns and employees by American Airlines.

TABLE OF CONTENTS

TABLE OF IMAGES

Terminology and Rules in the Airline Industry

Listed below are the definitions of commonly used terms in crew operations in the Airline Industry and some operational rules, relevant to my projects.

1. Crew: The crew on an airplane consists of the captain which is the main pilot, a co-pilot and/or a reserve pilot and the flight attendant staff.
2. Reserve flight-crew member: Reserve member applies to a flight crew member who is a certificate holder and is required to be available to receive an assignment for duty.

3. Lineholder: Lineholder means a flight crew member who has an assigned flight duty period and is not acting as a reserve flight crew member.

   The requirement of one or more than one reserve pilots depends on the type of airline equipment and the duration of a flight trip. For example, the Airbus 727 is a large sized aircraft capable to carry up to 500 passengers and is generally flown across longer trips. Hence, the flight crew of an Airbus 727 is inclusive of a co-pilot and a reserve pilot to accommodate for fatigue of the main pilot during longer flying hours.
   Pilots and reserves are allotted flight trips through the 'Preferential Bidding System'.
   A single flight crew consisting of the captain, copilot and reserves, is allowed to fly one type of airline equipment in a flying contract.

4. Crew Base: Every flight crew member is based out of a fixed airport city, which is the main resting location or residence of the crew members. The flight trips assigned to the crew members originate from their base locations.

5. Flight leg: Flight leg is the journey of an airplane from any airport (A) to an airport (B), from the time of take-off of the aircraft from origin airport until time of landing at the destination airport.

6. Flight Trip: A flight trip is defined as the journey of an aircraft and its crew from the base location (A) to another destination airport (B) in a single leg flight, or to multiple destinations (C) corresponding to the number of legs in a multi-legged flight and return back to the base location. (A)-(B), (B)-(C), (C)-(A).

7. Flight Duty Period: A duty period is the duration for which every crew member is responsible for performing his/her duties in flight. Generally, one duty period is calculated from the time of reporting of a crew member at the beginning of a flight leg until the member signs off for rest after landing. The total number of duty periods is equal to the number of legs in a flight trip.

8. Block: Block is the number of flying hours of a crew member during a trip. It is calculated as the duration from the time of takeoff from the origin airport to the time of landing at the destination airport.

9. Deadhead: If the last leg of the flight trip does not end at the base location, the flight crew is needed to fly the aircraft back to the base location without any passenger and this leg of the trip is called deadhead. To avoid unnecessary fuel costs without revenue, the last leg of the trip should ideally end at the crew base location. In addition, all time spent in deadhead transportation is duty and is not rest.

10. Calendar day: One calendar day implies to a 24-hour period from 0000 through 2359 using Coordinated Universal Time or local time.

11. Fatigue risk management system (FRMS): FRMS is a management system for a certificate holder to use to mitigate the effects of fatigue in its particular operations. It is a data-driven process and a systematic method used to continuously monitor and manage safety risks associated with fatigue-related error.

12. Rest over-night: The number of Rest Over-Nights (RONs) in a flight trip correspond to the number of nights spent by the crew away from the base location.

13. Synthetic: The Federal Aviation Administration (FAA) has imposed some limitations on the number of hours flown by the pilot and other crew members of an aircraft in one duty period of a flight trip, for purposes of passenger safety and well-being of the crew.
    The limitations have been made taking into consideration numerous different factors such as the number of hours spent away from the base, and the difference in time zones of the origin and destination, rest over nights, deadhead hours, etc.
    The crew members are provided extra compensation if the flight schedule of a trip does not meet the conditions imposed by FAA for any reason such as weather delays, maintenance issues, etc. The extra compensation awarded is determined by a set of complex equations having metrics from the factors contributing to the limitations stated above, as variables.

The salaries of the pilot and co-pilots of an aircraft and other crew members is a combination of hourly pay for the number of flying hours (block) and extra compensation awarded if the requirements set by FAA are not met.

'Synthetic' corresponds to the number of hours from the duty period for which extra compensation is provided to the crew members, excluding the hourly pay for the block.

14. Rig: The number of hours for which crew members are awarded extra compensation, excluding the deadhead hours, are called 'Rig'.

Thus, we have the following equations from the above definitions:

Duty Period = Block +Synthetic

Synthetic = Deadhead +Rig

Variance analysis of actual and planned synthetic and deadhead

Overview

The Crew performance and Operations analytics team at American Airlines is responsible for designing the block, synthetic, deadhead and other crew operations parameters. Crew operations expenses were at an all-time high for the business year 2015 and the first quarter Q1 of 2016. In the investigation of the causes for the rise in operational expenses, the management of the Crew strategy and Business Intelligence team was interested in visualizing the differences between the parameters planned by the Operations analytics team and these values that were actually observed. The visualization provided insights on the accuracy of parameters designed and helped the management identify the factors responsible for the variance. The project ultimately assisted stakeholders in targeting metrics with higher variability and develop a strategy to minimize variance and optimize operational costs.

Purpose and Scope of the project

The aim of the project was to visualize the variance between the actual and planned synthetic and deadhead for a period of 3 business quarters for flight trips starting from all the hubs for different aircraft equipment types. Along with synthetic and deadhead, the data for other operational performance metrics such as block hours, number of calendar days, and rest over nights was also extracted and analyzed. The project was initiated as a measure to visualize the deviation of metrics from the planned values and provide insights to the management to identify the causes of the variance.

The scope of the project included mining of planned and actual flight trip data for a period of three business quarters from multiple tables in a Teradata database by writing complex SQL queries; data preparation and exploratory data analysis in MS Excel and ultimately, presentation of results of the analysis using charts.

The project was executed in three phases –

1. Data mining of planned and actual flight trip data using T-SQL
2. Data preparation and validation in MS Excel
3. Exploratory data analysis and visualization

Phase 1: Data mining of planned and actual flight trip data using T-SQL

A. Extraction of Planned crew operational metrics

The flight trip data for metrics planned by the Crew Operations Analytics team corresponding to a period of 18 months – 01/01/2016 to 05/01/2017 was extracted by writing the following nested query on T-SQL.

```
select b.yr, b.m, b.ttl_dp, b.ttl_cal_day, b.long_ron, b.short_ron, b.ttl_ron,
        extract(day from b.ttl_blk) * 24 * 60 + extract(hour from b.ttl_blk) *60 + extract(minute from b.ttl_blk) as ttl_blk,
        extract(day from b.ttl_dh) * 24  * 60+ extract(hour from b.ttl_dh) *60 + extract(minute from b.ttl_blk) as ttl_dh,
        extract(day from b.ttl_syn) * 24  * 60+ extract(hour from b.ttl_syn) *60 + extract(minute from b.ttl_blk) as ttl_syn
from
(
    select
            FLIGHT_CREW_CONTRCT_YEAR as yr,
            FLIGHT_CREW_CONTRCT_Month as m,
            sum(a.SEQ_DUTY_PERIOD_CT) as ttl_dp,
            sum(a.SEQ_CALNDR_DAY_CT) as ttl_cal_day,
            sum(a.LONG_RON_CT) as long_ron,
            sum(a.SHORT_RON_CT) as short_ron,
            sum(a.TTL_RON_CT) as ttl_ron ,
            sum(a.TTL_BLOCK_TM) as ttl_blk,
            sum(a.TTL_DEDHED_TM) as ttl_dh,
            sum(a.DUTY_RIG_TM) as ttl_syn
    from
    (
        SELECT  distinct FLIGHT_CREW_CONTRCT_YEAR, FLIGHT_CREW_CONTRCT_MONTH,
                seq_nbr, seq_schd_start_dt,
                SEQ_DUTY_PERIOD_CT,DUTY_RIG_TM,
                SEQ_CALNDR_DAY_CT, DEDHED_FLIGHT_LEG_CT,
                LONG_RON_CT, SHORT_RON_CT, TTL_RON_CT, TTL_BLOCK_TM, TTL_DEDHED_TM

        FROM    PROD_FLIGHT_CREW_VWS.FLIGHT_CREW_PAIRED_SEQ
        where ROW_EXPIRY_DT ='9999-12-31' and SEQ_CREW_POSITN_GROUP_CD = 'fc'
        and SEQ_CREW_POSITN_CT = 2 and FLIGHT_CREW_GROUP_CD = 'pi'
        and seq_schd_start_dt between '2016-01-01' and '2017-05-01'
        and AIRCFT_EQUIP_FLEET_CD = '737'
    ) a

    group by yr, m
) b
order by 1,2
```

Figure. 2. Snapshot of T-SQL Query for extraction of planned crew operations metrics

The table used for extraction of planned flight trip data was ‘FLIGHT_CREW_PAIRED_SEQ’ which was stored in the relational database view or virtual table called ‘PROD_FLIGHT_CREW_VWS’.

From the FLIGHT_CREW_PAIRED_SEQ table, the columns listed below were selected in the inner loop of the query. The explanation of the data in these columns is also stated below:

FLIGHT_CREW_CONTRCT_YEAR: Contract year of crew members

FLIGHT_CREW_CONTRCT_MONTH: Contract month of crew members

seq_nbr: Unique identifier of every flight trip

seq_schd_start_dt: Start date of a flight trip

SEQ_DUTY_PERIOD_CT: Number of duty periods in a trip

DUTY_RIG_TM: Rig time in a flight trip in hours

SEQ_CALNDR_DAY_CT: Duration of the flight trip in days

DEDHED_FLIGHT_LEG_CT: Number of deadhead legs in a trip

LONG_RON_CT: Number of long rest over nights in a trip

SHORT_RON_CT: Number of short rest over nights in a trip

TTL_RON_CT: Total number of rest over nights in a trip

TTL_BLOCK_TM: Total block time in a trip

TTL_DEDHED_TM: Total deadhead time in a trip

The where clauses used in the inner loop of the query correspond to the following columns:

ROW_EXPIRY_DT: Indicator for showing the validity of a trip. Set to ‘9999-12-31’ (invalid date) for all rows with valid sequence numbers in the FLIGHT_CREW_PAIRED_SEQ table, indicating that the trip sequence is planned and has not taken place yet.

SEQ_CREW_POSITN_GROUP_CD: Indicator for the crew group code. Set to 'fc' which means ‘flight crew’.

SEQ_CREW_POSITN_CT = 2 and FLIGHT_CREW_GROUP_CD = 'pi': Both these parameters specify the type of crew member for which the data is extracted. Crew group code ‘pi’ and crew position count = 2, both mean that the data corresponds to crew member ‘Pilot’.

seq_schd_start_dt between '2016-01-01' and '2017-05-01': Start date of the flight trip sequence lies between ‘2016-01-01’ and ‘2017-05-01’

AIRCFT_EQUIP_FLEET_CD = '737': Indicates that the aircraft equipment type used for the flight trip has the fleet code ‘737’ which corresponds to the aircraft Boeing 737.

As the interest of the project lies in the calculation and comparison of the operational parameters as single entities, we create variables called ttl_dp, ttl_cal_day, long_ron, short_ron, ttl_ron, ttl_blk, ttl_dh, and ttl_syn representing the total number of hours of duty period, total number of calendar days, total long rons, total short rons, total block and total synthetic for all pilots flying the aircraft type Boeing 737 between the dates ‘2016-01-01' and '2017-05-01'. The ‘sum’ function is used in the second layer of the query for adding up all the metric values.

The total block, synthetic and deadhead are originally stored in the ‘dd:hr:mm’ (Days, hours, minutes) format while the requirements of the project demands the representation of these parameters in hours, to facilitate ease of comparison; as the compensation provided to the crew members is on an hourly basis. Therefore, in the outermost layer of the query, number of hours in the total block, synthetic and deadhead are extracted.

By making changes in the where clauses in the above query, the planned flight trip data for crew member other than pilots, such as reserves or flight attendants, or trips of different aircraft equipment types can be extracted by changing the respective codes. To restrict the information extracted for flight trip data to only the trips originating from a certain base airport location, an additional where clause can be included in the query, by assigning the base location code of the origin airport to the variable FLIGHT_CREW_BASE_CD.

The extracted data for planned flight trips from the above query is shown below in Figure 3. Columns ‘yr’ and ‘m’ specify the year and month.

	A	B	C	D	E	F	G	H	I	J
1	yr	m	ttl_dp	ttl_cal_day	long_ron	short_ron	ttl_ron	ttl_blk	ttl_dh	ttl_syn
2	2,016	1	2,756	2,779	485	899	1,384	958,396	14,236	25,876
3	2,016	2	2,811	2,824	467	943	1,410	982,111	16,171	30,871
4	2,016	3	2,981	3,010	453	1,102	1,555	1,005,173	22,973	40,973
5	2,016	4	2,946	2,948	408	1,231	1,639	1,021,003	23,563	33,343
6	2,016	5	2,939	2,951	541	1,012	1,553	1,022,120	21,920	39,560
7	2,016	6	3,047	3,157	571	1,033	1,604	1,044,013	24,853	47,413
8	2,016	7	3,086	3,209	501	1,154	1,655	1,052,849	21,209	44,129
9	2,016	8	3,103	3,190	605	1,065	1,670	1,041,292	22,252	48,232
10	2,016	9	2,967	3,008	546	1,079	1,625	984,276	20,976	40,476
11	2,016	10	3,141	3,204	677	997	1,674	1,004,666	23,666	48,506
12	2,016	11	3,150	3,225	696	984	1,680	1,004,664	25,704	54,564
13	2,016	12	3,236	3,326	710	1,058	1,768	1,030,412	23,972	55,772
14	2,017	1	3,040	3,102	696	1,001	1,697	998,510	17,450	47,330
15	2,017	2	3,030	3,103	647	1,023	1,670	1,000,378	16,078	46,438
16	2,017	3	3,344	3,427	698	1,179	1,877	1,090,757	17,717	52,517
17	2,017	4	3,260	3,330	647	1,139	1,786	1,077,895	11,515	40,855
18	2,017	5	3,275	3,407	679	1,173	1,852	1,103,776	15,316	42,496
19										

Figure 3: Snapshot of planned flight trip data

B. Extraction of actual crew operational metrics

The actual flight trip data was extracted from the table FLIGHT_CREW_EMP_SEQ present in the virtual table 'PROD_FLIGHT_CREW_VWS' in a manner similar to the extraction of planned flight trip data.

The relevant column names in the FLIGHT_CREW_EMP_SEQ table were the same as those in the 'FLIGHT_CREW_PAIRED_SEQ' table used for planned data except that the column for aircraft equipment type code was not present in the FLIGHT_CREW_EMP_SEQ table.

It was necessary to segregate the data based on the type of aircraft equipment and hence the table in use was joined with another table FLIGHT_CREW_EMP_ASGN_FLIGHT in the database which contained the column 'SCHD_AIRCRFT_EQUIP_TYPE_CD' referring to the type of aircraft equipment.

A left join is performed in the query to retain all entries from the FLIGHT_CREW_EMP_SEQ table corresponding to the actual flight trip data with the flight trip sequence number column 'seq_nbr' as the common entity between the two tables. The aircraft equipment type code '737' is specified in the where clause following the left join.

```
 select b.yr, b.m, b.ttl_dp, b.ttl_cal_day, b.long_ron, b.short_ron, b.ttl_ron,
                          extract(day from b.ttl_blk) * 24 * 60 + extract(hour from b.ttl_blk) *60 + extract(minute from b.ttl_blk) as ttl_blk,
                          extract(day from b.ttl_dh) * 24  * 60+ extract(hour from b.ttl_dh) *60 + extract(minute from b.ttl_blk) as ttl_dh,
                          extract(day from b.ttl_syn) * 24  * 60+ extract(hour from b.ttl_syn) *60 + extract(minute from b.ttl_blk) as ttl_syn
from
(
        select
                                  FLIGHT_CREW_CONTRCT_YEAR as yr,
                                  FLIGHT_CREW_CONTRCT_Month as m,
                                  sum(a.SEQ_DUTY_PERIOD_CT) as ttl_dp,
                                  sum(a.SEQ_CALNDR_DAY_CT) as ttl_cal_day,
                                  sum(a.LONG_RON_CT) as long_ron,
                                  sum(a.SHORT_RON_CT) as short_ron,
                                  sum(a.TTL_RON_CT) as ttl_ron ,
                                  sum(a.TTL_BLOCK_TM) as ttl_blk,
                                  sum(a.TTL_DEDHED_TM) as ttl_dh,
                                  sum(a.DUTY_RIG_TM) as ttl_syn
        from
        (
                SELECT  distinct FLIGHT_CREW_CONTRCT_YEAR, FLIGHT_CREW_CONTRCT_MONTH,
                                seq_nbr, seq_schd_start_dt,
                                SEQ_DUTY_PERIOD_CT,DUTY_RIG_TM,
                                SEQ_CALNDR_DAY_CT, DEDHED_FLIGHT_LEG_CT,
                                LONG_RON_CT, SHORT_RON_CT, TTL_RON_CT, TTL_BLOCK_TM, TTL_DEDHED_TM

                FROM    PROD_FLIGHT_CREW_VWS.FLIGHT_CREW_EMP_SEQ
                where ROW_EXPIRY_DT ='9999-12-31' and SEQ_CREW_POSITN_GROUP_CD = 'fc'
                and SEQ_CREW_POSITN_CT = 2 and FLIGHT_CREW_GROUP_CD = 'pi'
                and seq_schd_start_dt between '2016-01-01' and '2017-05-01'
        ) a

        left join PROD_FLIGHT_CREW_VWS.FLIGHT_CREW_EMP_ASGN_FLIGHT p
        on a.seq_nbr = p.seq_nbr
        where p.SCHD_AIRCRFT_EQUIP_TYPE_CD = '737'

        group by yr, m
) b
order by 1,2
```

Figure 4: Snapshot of T-SQL query to extract actual crew operational metrics

The highlighted portion of the query in the above image shows the left join on the FLIGHT_CREW_EMP_ASGN_FLIGHT table and the where clause to specify the aircraft Boeing 737

The actual flight trip data extracted from the above query was as follows in Figure 5.

	A	B	C	D	E	F	G	H	I	J
1	FLIGHT_CREW_CONTRCT_YEAR	FLIGHT_CREW_CONTRCT_MONTH	EMP_NUM	SEQ_SCHD_START_DT	DP	TTL_CLNDR_DY	BLOCK	DEADHEAD	SYNTHETIC	RONS
2	2,016	1	9,591	1/20/2016	1	1	69	67	164	0
3	2,016	1	9,591	1/9/2016	1	1	488	0	0	0
4	2,016	1	9,591	1/21/2016	1	1	427	0	0	0
5	2,016	1	9,591	1/1/2016	1	1	416	0	0	0
6	2,016	1	9,591	1/18/2016	1	1	420	0	0	0
7	2,016	1	9,591	1/6/2016	1	1	430	0	0	0
8	2,016	1	9,591	1/3/2016	1	1	436	0	0	0
9	2,016	1	9,628	1/19/2016	1	1	492	0	0	0
10	2,016	1	9,628	1/10/2016	1	1	430	0	0	0
11	2,016	1	9,628	1/25/2016	1	1	430	0	0	0
12	2,016	1	9,628	1/18/2016	1	1	422	0	0	0
13	2,016	1	9,628	1/17/2016	1	1	461	0	0	0
14	2,016	1	9,628	1/6/2016	1	1	499	0	0	0
15	2,016	1	9,628	1/24/2016	1	1	415	0	0	0
16	2,016	1	9,628	1/1/2016	1	1	352	0	0	0
17	2,016	1	9,628	1/5/2016	1	1	471	0	0	0
18	2,016	1	9,628	1/26/2016	1	1	406	0	0	0
19	2,016	1	9,652	1/7/2016	1	1	339	0	0	0
20	2,016	1	9,652	1/2/2016	1	1	425	0	0	0
21	2,016	1	9,652	1/19/2016	1	1	356	0	0	0

Figure 5: Snapshot of Actual flight trip data

The data was cleaned by removing duplicate values by sorting the data first by employee numbers and then using the ‘remove duplicates’ option in Excel.

Phase 2: Data validation and preparation

The next step of the project was validating the extracted data against another data source and preparing the data for analysis.

The actual flight trip data extracted as stated above was validated against the mainframe data dashboard which is a live data source of all the flight trips information stored year-to-date. By using filters on the dates of departure and aircraft equipment type on the mainframe dashboard, corresponding flight trip information was extracted for comparison. The total number of flight trips were checked against the trips from mainframe using vlookups, index match and other Advanced Excel tools using the flight trip sequence number columns in both datasets.

The sequences not matching in any one of the datasets as well as missing values were investigated.

It was observed that the number of sequences occurring in the actual data extracted, was higher than the number of cases in the mainframe dashboard by about 60 sequences. Upon detailed investigation, it was revealed that one of the parameters in the actual dataset 'TRIP_SELECT_TYPE_CD' could take any one the following set of values: (0, 1, 2, 3, 4). The values (0, 2, 3, 4) for the 'TRIP_SELECT_TYPE_CD' meant that the flight crew member in consideration were 'Lineholders'- pilots and co-pilots; whereas if the 'TRIP_SELECT_TYPE_CD' had the value 1, meant that the information was for reserve pilots or co-pilots who had been called for, for certain safety reasons, but not actually placed on duty. Since the mainframe data only showed the crew members actually placed on duty, the rows corresponding to 'TRIP_SELECT_TYPE_CD' = '1' had to be excluded from the actual flight trip dataset. Hence, the required changes were made in the query for actual data by adding a where clause for the column 'TRIP_SELECT_TYPE_CD'. The inclusion of this where clause significantly reduced the number of non-matching flight trip sequences thus improving the accuracy of the data extracted.

The data was again verified against another dataset called the ‘Utilization Data’ which corresponded to the actual hourly payment information of the crew members. On finding very few of missing values and outliers after the validation with the ‘Utilization Data’, the integrity of the extracted actual dataset was verified.

Shown below in figure 6 is a snapshot of the ‘Utilization data’ extracted for the purpose of validation.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	EMP_NBR	FLIGHT_CRI	FLIGHT_CRI	FLIGHT_CRI	DIVISION	BASE	SEAT_POS	CREW_EQU	TRIP_TYPE	HEADCNT_	TOT_MIN	Productive	SYNTHETIC	DEADHEAD	UNDERFLY	Block
2	9591	2016	8	PI	D	DFW	CA	737	Line Holder	1	4413	4048	5	0	105	3938
3	9591	2016	9	PI	D	DFW	CA	737	Line Holder	1	4207	3770	0	0	151	3619
4	9591	2016	2	PI	D	DFW	CA	737	Line Holder	1	5134	0	0	0	0	0
5	9591	2017	4	PI	D	DFW	CA	737	Line Holder	1	4684	2744	0	0	125	2619
6	9591	2016	1	PI	D	DFW	CA	737	Line Holder	1	5445	3023	164	67	106	2686
7	9591	2016	4	PI	D	DFW	CA	737	Line Holder	1	5322	0	0	0	0	0
8	9591	2017	1	PI	D	DFW	CA	737	Line Holder	1	5226	3914	0	0	125	3789
9	9591	2016	5	PI	D	DFW	CA	737	Line Holder	1	4747	2524	0	0	58	2466
10	9591	2017	3	PI	D	DFW	CA	737	Line Holder	1	4282	3040	0	0	140	2900
11	9591	2016	12	PI	D	DFW	CA	737	Line Holder	1	5002	358	0	0	27	331
12	9591	2016	10	PI	D	DFW	CA	737	Line Holder	1	4528	3826	33	0	141	3652
13	9591	2016	7	PI	D	DFW	CA	737	Line Holder	1	3391	2305	0	0	68	2237
14	9591	2017	2	PI	D	DFW	CA	737	Line Holder	1	5122	4752	7	150	190	4405
15	9591	2016	3	PI	D	DFW	CA	737	Line Holder	1	5086	0	0	0	0	0
16	9591	2016	6	PI	D	DFW	CA	737	Line Holder	1	4810	4381	0	0	113	4268
17	9591	2016	11	PI	D	DFW	CA	737	Line Holder	1	4537	2502	0	0	57	2445
18	9628	2016	2	PI	D	DFW	CA	737	Line Holder	1	5661	4639	0	254	86	4299
19	9628	2017	3	PI	D	DFW	CA	737	Line Holder	1	4285	2027	23	0	66	1938
20	9628	2017	1	PI	D	DFW	CA	737	Reserves	0.03	243	0	0	0	0	0
21	9628	2016	5	PI	D	DFW	CA	737	Line Holder	1	5468	5224	0	0	117	5107
22	9628	2016	10	PI	D	DFW	CA	737	Line Holder	1	4532	0	0	0	0	0
23	9628	2016	3	PI	D	DFW	CA	737	Line Holder	1	5849	4396	0	0	45	4351
24	9628	2016	1	PI	D	DFW	CA	737	Line Holder	1	5293	4437	0	0	59	4378
25	9628	2016	8	PI	D	DFW	CA	737	Line Holder	1	5046	3111	0	0	80	3031
26	9628	2016	9	PI	D	DFW	CA	737	Line Holder	1	4875	0	0	0	0	0
27	9628	2016	4	PI	D	DFW	CA	737	Line Holder	1	6016	5836	0	0	235	5601
28	9628	2017	4	PI	D	DFW	CA	737	Line Holder	1	5713	5005	24	0	219	4762
29	9628	2016	11	PI	D	DFW	CA	737	Line Holder	1	4411	0	0	0	0	0
30	9628	2016	6	PI	D	DFW	CA	737	Line Holder	1	4884	4884	0	0	212	4672

Figure 6: Snapshot of flight utilization data for validation

Phase 3: Exploratory data analysis and visualization of actual and planned crew operational metrics

After the data validation process, the two clean datasets for actual and planned flight trip were prepared for analysis. The two datasets were imported into one Excel worksheet along with the cleaned dataset of the Utilization Data.

The following chart (Figure 7) was prepared based on the monthly sum of the values of synthetic and deadhead from the planned, actual and utilization datasets for the year 2017. The 'sumif' function was used to calculate the total values of the crew operations metrics from the datasets. Variance between Actual values and planned (scheduled) metric values, as well as between Actual and Utilization values was calculated.

Equipment	737
Position	CA
Base	DFW

				2017		
		Jan	Feb	March	April	May
	Scheduled	788.8	774.0	875.3	680.9	708.3
Synthetic	Actual	1231.0	1218.4	1277.1	1332.5	1374.2
	Utilization Data	1171.2	1207.7	1257.5	1270.0	1334.1
	Variance (Actl, Schd)	442.2	444.4	401.8	651.6	665.9
	Variance (Actl,Utl)	59.8	10.7	19.6	62.5	40.0
	Scheduled	290.8	268.0	295.3	191.9	255.3
Deadhead	Actual	1137.5	915.0	1087.3	1034.0	1021.3
	Utilization Data	1069.8	879.0	1047.5	968.9	953.1
	Variance (Actl, Schd)	846.7	647.0	792.0	842.0	766.0
	Variance (Actl,Utl)	67.8	36.0	39.9	65.1	68.2

Figure 7: Chart showing monthly values of operational metrics for months January to May 2017

The above chart (Figure 7) shows the crew operational metrics information for all captains of aircraft Boeing 737, starting from the base airport Dallas Fort Worth (DFW). The chart could be filtered by a drop-down list for different equipment types, position of the crew members such as captain (CA), co-pilots (CP) and/or reserves (R) as well as different base airport locations.

The bar chart below in Figure 8 shows the graphical representation of Total Synthetic Hours of Scheduled (Planned), Actual and Utilization Data. It also represents the variance between Actual and Scheduled and Actual and Utilization Data.

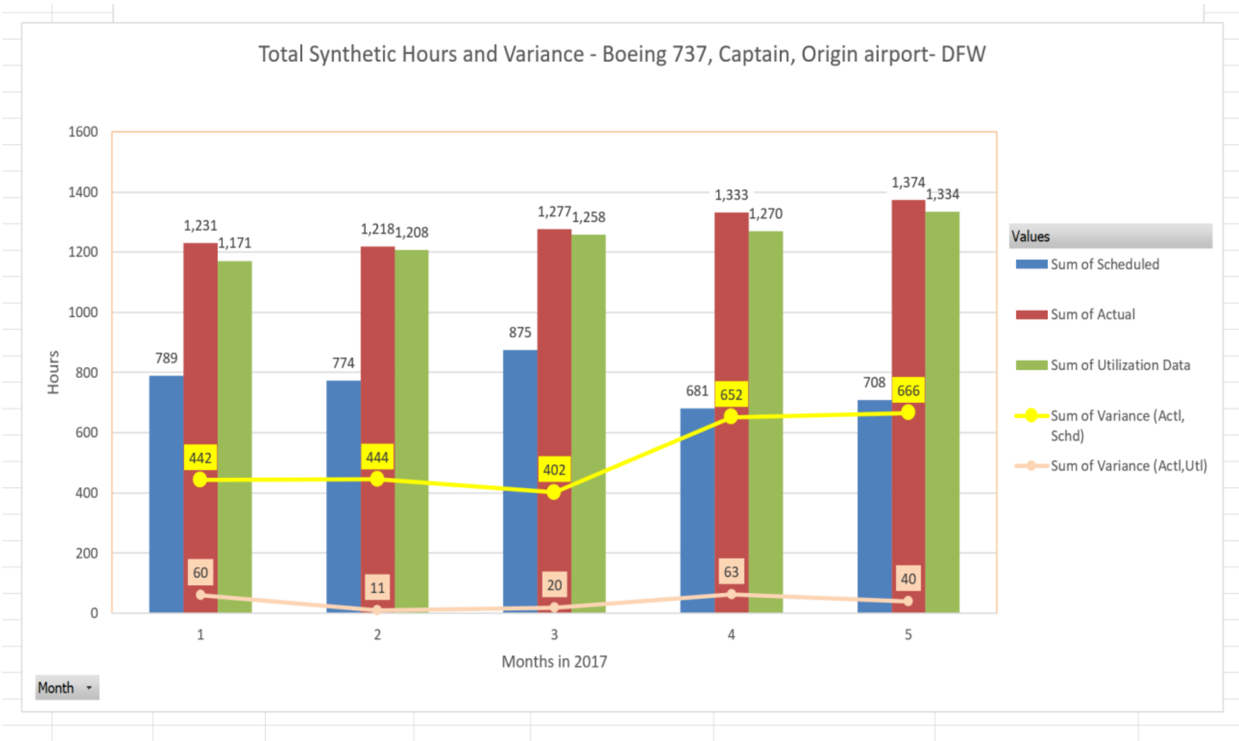

Figure 8: Total Synthetic Hours and Variance for Boeing 737, Captain, at origin airport DFW

The bar chart below in Figure 9 shows the graphical representation of Total Deadhead Hours of Scheduled (Planned), Actual and Utilization Data. It also represents the variance between Actual and Scheduled and Actual and Utilization Data.

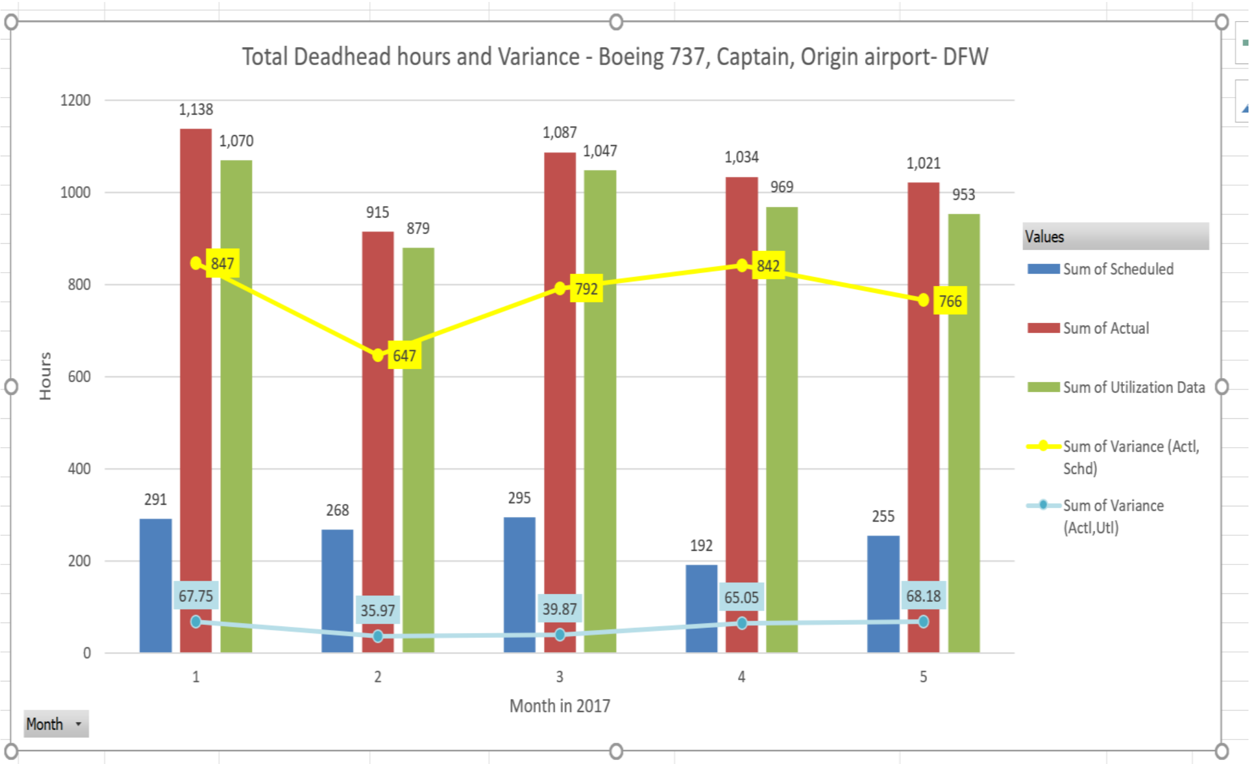

Figure 9: *Total Deadhead Hours and Variance for Boeing 737, Captain, at origin airport DFW*

Results and Conclusion

From the above analyses and visualizations, it was clear that the total variance between actual and planned synthetic was in the range of 442- 666 hours in the first 5 months of 2017, with an average of 554 hours.

The difference between actual and planned deadhead hours was in the range of 747-847 hours with an average of 787 hours. This variance is significant and relates to the need for improvement in accuracy of the forecasts for the operations metrics. It directly corresponds to the increase in operational costs occurred in the year 2017 and suggests a need for significant improvement in efficiency with respect to flight trip operational activities.

There is also some amount of variability in the actual and utilization data as observed from the results of analysis. The cause for this variability also needs to be investigated and points towards an issue of data integrity and inhomogeneity within the data infrastructure of the organization and calls for a need to asses and analyze the data sources and entry methods of the system.

The results of the analysis provided valuable insights to the management regarding issues within the organization and provided recommendation about areas needing improvement.

Future Scope

The future scope of the project included improving the accuracy of analysis and inclusion of more factors like trip start time, seasonal effects and air traffic conditions within the data model for improving the viability of the analytical model.

The results of the analysis could be used to formulate a strategy roadmap by providing data-based solution for targeting the issue areas and thus improving the overall functional efficiency of crew operations.

Bibliography

Electronic code of federal regulations (e- CFR)

https://www.aa.com/i18n/customer-service/about-us/history-of-american-airlines.jsp

Mosiac Intermediate SQL

https://www.sas.com/en_us/learn/academic-programs/.../free-sas-e-learning.html

Appendix

Source: https://www.faa.gov/regulations_policies/handbooks_manuals/

1. Cumulative limitations imposed by Federal Aviation Administration (FAA):

No flight crew member may accept an assignment if the crew member's total flight time will exceed the following:

(1) 100 hours in any 672 consecutive hours or

(2) 1,000 hours in any 365 consecutive calendar day period.

No flight crew member may accept an assignment if the crew member's total Flight Duty Period will exceed:

(1) 60 flight duty period hours in any 168 consecutive hours or

(2) 190 flight duty period hours in any 672 consecutive hours

2. Fatigue risk management system:

FAA under a Fatigue Risk Management System provides for a level of safety against fatigue-related accidents or incidents as the other provisions of this part.

The Fatigue Risk Management System must include:

(1) A fatigue risk management policy.

(2) An education and awareness training program.

(3) A fatigue reporting system.

(4) A system for monitoring flight crew fatigue.

(5) An incident reporting process.

(6) A performance evaluation

3. Flight time limitation:

No crew member may accept an assignment or continue an assigned flight duty period if the total flight time:

(1) Will exceed the limits specified in Table A of this part if the operation is conducted with the minimum required flight crew.

(2) Will exceed 13 hours if the operation is conducted with a 3-pilot flight crew.

(3) Will exceed 17 hours if the operation is conducted with a 4-pilot flight crew.

4. Flight duty period: Augmented flight crew:

(a) For flight operations conducted with an acclimated augmented flight crew, no flight crew member may accept an assignment if the scheduled flight duty period will exceed the limits specified in Table C displayed later.

(b) If the flight crew member is not acclimated:

(1) The maximum flight duty period in Table C of this part is reduced by 30 minutes.

(2) The applicable flight duty period is based on the local time at the theater in which the flight crew member was last acclimated.

(c) No flight crew member may accept an assignment unless during the flight duty period:

(1) Two consecutive hours in the second half of the flight duty period are available for in-flight rest for the pilot flying the aircraft during landing.

(2) Ninety consecutive minutes are available for in-flight rest for the pilot performing monitoring duties during landing.

Below are the tables for Flight time limits for Augmented and Unaugmented Operations:

Table A to Part 117—Maximum Flight Time Limits for Unaugmented Operations Table

Time of report (acclimated)	Maximum flight time (hours)
0000-0459	8
0500-1959	9
2000-2359	8

Table B to Part 117—Flight Duty Period: Unaugmented Operations

Scheduled time of start (acclimated time)	Maximum flight duty period (hours) for lineholders based on number of flight segments						
	1	2	3	4	5	6	7 +
0000-0359	9	9	9	9	9	9	9
0400-0459	10	10	10	10	9	9	9
0500-0559	12	12	12	12	11.5	11	10.5
0600-0659	13	13	12	12	11.5	11	10.5
0700-1159	14	14	13	13	12.5	12	11.5
1200-1259	13	13	13	13	12.5	12	11.5
1300-1659	12	12	12	12	11.5	11	10.5
1700-2159	12	12	11	11	10	9	9
2200-2259	11	11	10	10	9	9	9
2300-2359	10	10	10	9	9	9	9

Table C to Part 117—Flight Duty Period: Augmented Operations

Scheduled time of start (acclimated time)	Maximum flight duty period (hours) based on rest facility and number of pilots					
	Class 1 rest facility		Class 2 rest facility		Class 3 rest facility	
	3 pilots	4 pilots	3 pilots	4 pilots	3 pilots	4 pilots
0000-0559	15	17	14	15.5	13	13.5
0600-0659	16	18.5	15	16.5	14	14.5
0700-1259	17	19	16.5	18	15	15.5
1300-1659	16	18.5	15	16.5	14	14.5
1700-2359	15	17	14	15.5	13	13.5