TABLEAU PROJECT REPORT

Paddy Peng WANG

INSTRUCTOR: MR. HAMID RAJAEE

Contents

Flights Delay Analysis

USING TABLEAU

BY PENG WANG

Metro College of Technology
Instructor: Mr. Hamid Rajaee

METRO COLLEGE OF TECHNOLOGY

ON Time? NO Time?

How long have you not travelled with your flight on time?

According to the Bureau of Statistics, about 20% of all flights are delayed by 15 minutes or more[1]. During ClaimCompass airline ratings case study, that figure is closer to 24.3%[2].

Flights delay have negative impact on passengers, airlines, and economic. In particular, flight delays severely decrease customer satisfaction and might lead to customers choosing a different airport or airline, in the long term.

As a passenger, an important benchmark for evaluating travel options, e.g. in terms of airports, airlines or even modes of transportation (train vs plane) is the punctuality of each option.

Customer satisfaction with airlines and airport performance is considerably influenced by how much flights are delayed. But how should the delay be quantified with thousands of flights for each airport and airline? Here, I present a statistical analysis of departures at several New York City airports in 2013.

Understand Data

The dataset that I collected contains:

A. Flights departure time details from three major NYC airports, namely:

- Newark Liberty International Airport (EWR),
- John F. Kennedy International Airport (JFK),
- LaGuardia Airport (LGA).

B. Flights arrival time details from 105 destination airports around the world.

For a full year of 2013, the data combined 16 airline companies operating the flights along with the corresponding flight number, total of approximately three-hundred and thirty-six thousand (336,776) records of 4000 flights departures and arrivals details related with the airports mentioned above.

The delay is then computed simply as the difference between an aircraft's scheduled departure/arrival time and its actual departure/arrival time.

For example:

Departure Delay = Departure Time – Scheduled Departure Time

- If Departure Delay > 0 ➔ Delayed
- If Departure Delay = 0 ➔ On-Time
- If Departure Delay < 0 ➔ Ahead

The same logic applied to Arrival Delay.

Based on the Departure Delay and Arrival Delay, I conducted the relationship analysis between these two factors, also other factors that might impact the flights delay (or ahead) results.

Business Requirement

There are many reasons caused a flight delay or ahead, such as weather, natural disaster, air traffic control, strikes, or even a bird strike!

However, for this project, based on the available data features, I focused on below factors and their relationship with flights punctual performance.

- Departure Status (Ahead/Delayed/On-time)
- Arrival Status (Ahead/Delayed/On-time)
- Carrier (16 companies)
- Origin (Departure) Airport (EWR/JFK/LGA)

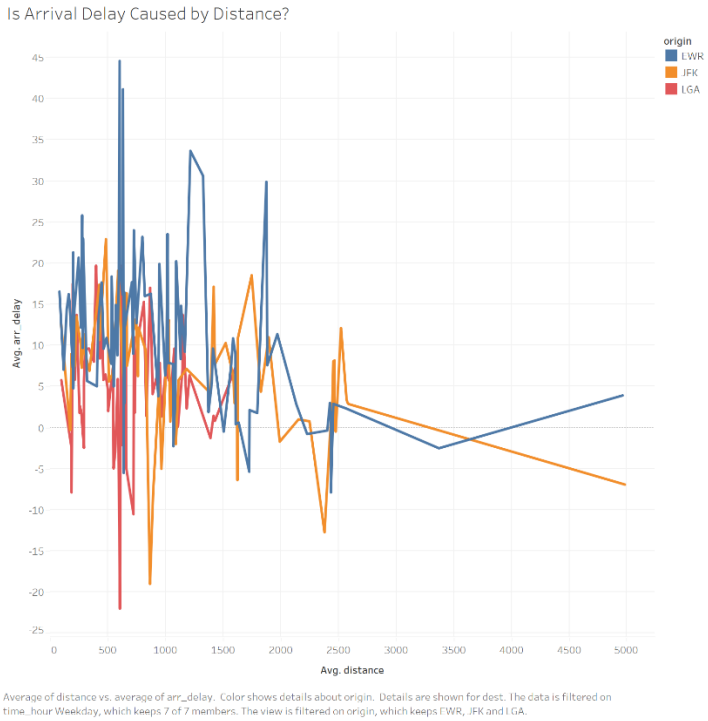
- Distance (in KM)
- Departure Hour (5-23)
- Departure Weekday (Sun-Sat)

My goal is to answer below business questions:

1. **Does departure delay always cause arrival delay?**
2. **Which airline(carrier) has the best/worst punctual performance?**
3. **Which day or hour shall we pay attention to control delay better?**

Some Interesting Findings

Distance never causes an arrival delay.



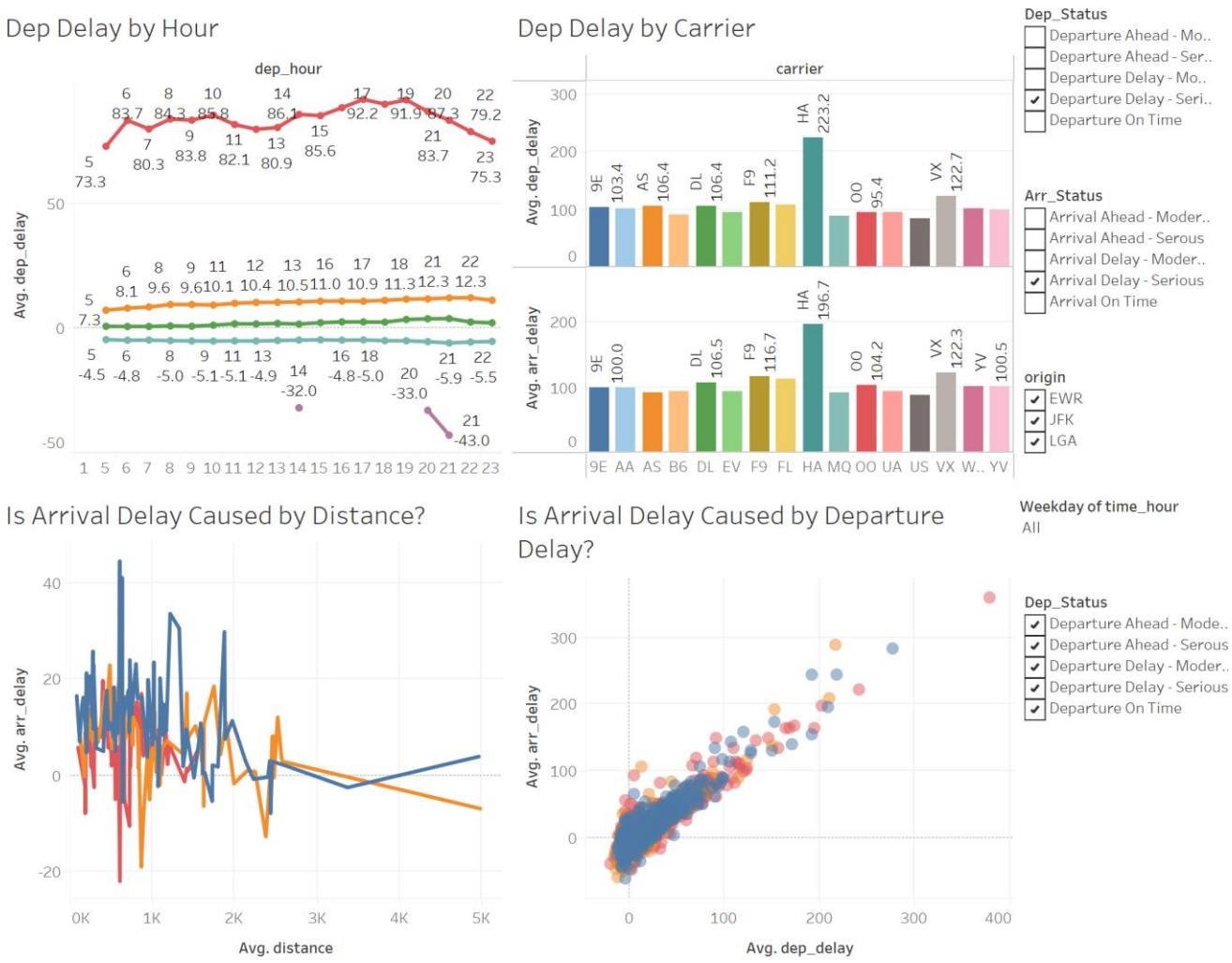
EWR – the most troublesome but not the worst airport.

If just by looking at the number of flights had serious departure delay (more than 30 minutes), then obviously EWR will be the worst airport with 19914 flights had been seriously delayed.

However, if we look the average delay time among the three airports, EWR is the best one with the lowest 85.66 minutes delay. While it's still not in an ideal situation but compare with other two airports, it has the shortest serious delay time.

Sum of dep_delay and average of dep_delay for each Dep_Status broken down by origin. Color shows details about origin. The marks are labeled by origin. For pane Sum of dep_delay: The marks are labeled by origin and count of flight. For pane Average of dep_delay: The marks are labeled by origin and average of dep_delay. The data is filtered on time_hour Weekday, which keeps 7 of 7 members. The view is filtered on Dep_Status and origin. The Dep_Status filter keeps Departure Delay - Moderate and Departure Delay - Serious. The origin filter keeps EWR, JFK and LGA.

Conclusions



Let us go back to the original business questions:

1. Does departure delay always cause arrival delay?

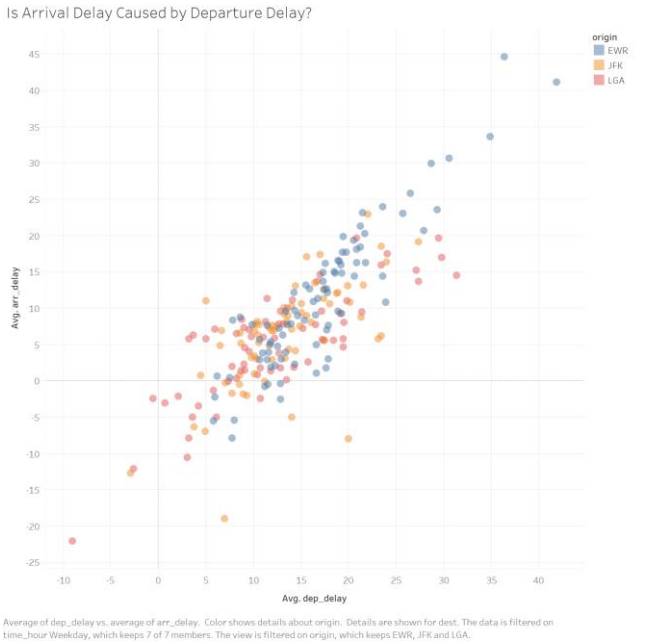
2. Which airline(carrier) has the best/worst punctual performance?

3. Which day or hour shall we pay attention to control delay better?

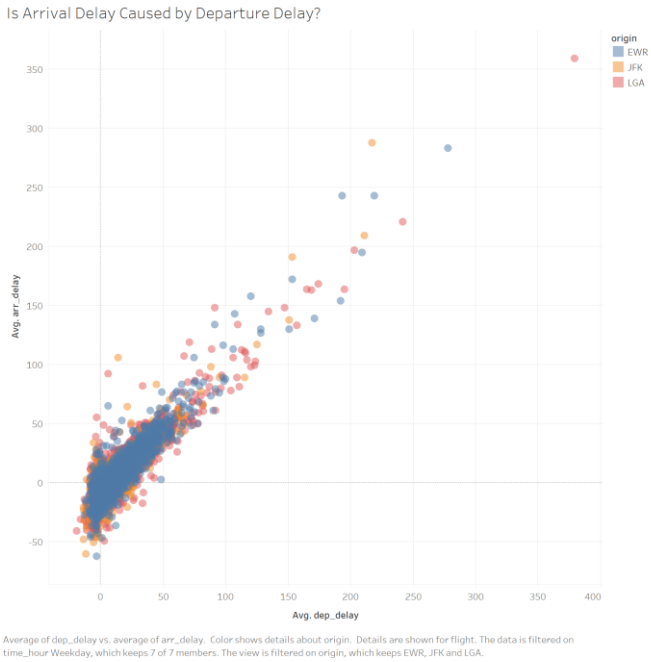
Based on the analysis, we have below conclusions:

1. Depart delay has positive impact on arrival delay.

Review the relationship by each destination:

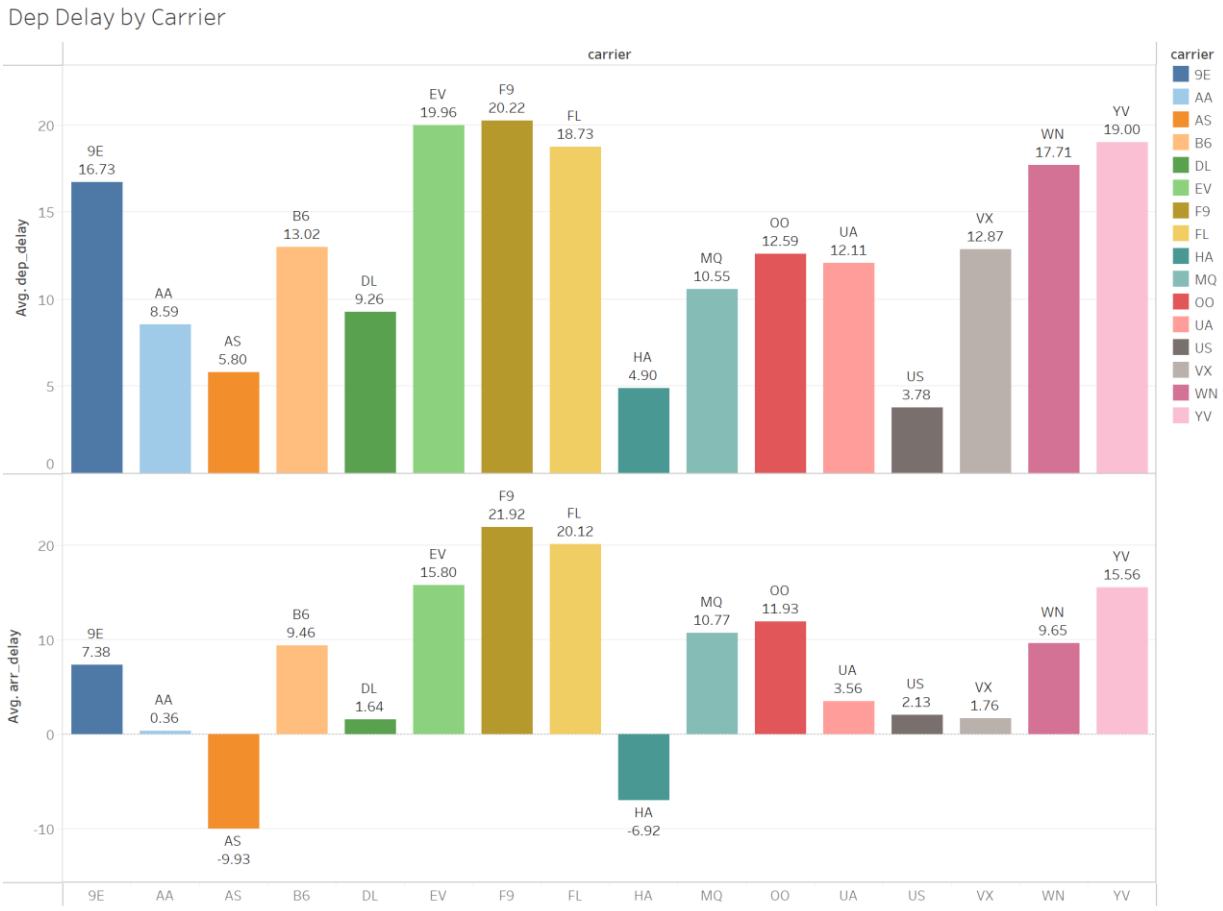


Or, by each flight:



It seems obvious that the more delayed on departure, the more delayed on arrival.

2. Carrier F9 has the worst punctual issue while US airline is the best.



Average of dep_delay and average of arr_delay for each carrier. Color shows details about carrier. The marks are labeled by carrier. For pane Average of dep_delay: The marks are labeled by carrier and average of dep_delay. For pane Average of arr_delay: The marks are labeled by carrier and average of arr_delay. The data is filtered on Dep_Status, Arr_Status, time_hour Weekday and origin. The Dep_Status filter keeps Departure Ahead - Moderate, Departure Ahead - Serous, Departure Delay - Moderate, Departure Delay - Serious and Departure On Time. The Arr_Status filter keeps Arrival Ahead - Moderate, Arrival Ahead - Serous, Arrival Delay - Moderate, Arrival Delay - Serious and Arrival On Time. The time_hour Weekday filter keeps 7 of 7 members. The origin filter keeps EWR, JFK and LGA.

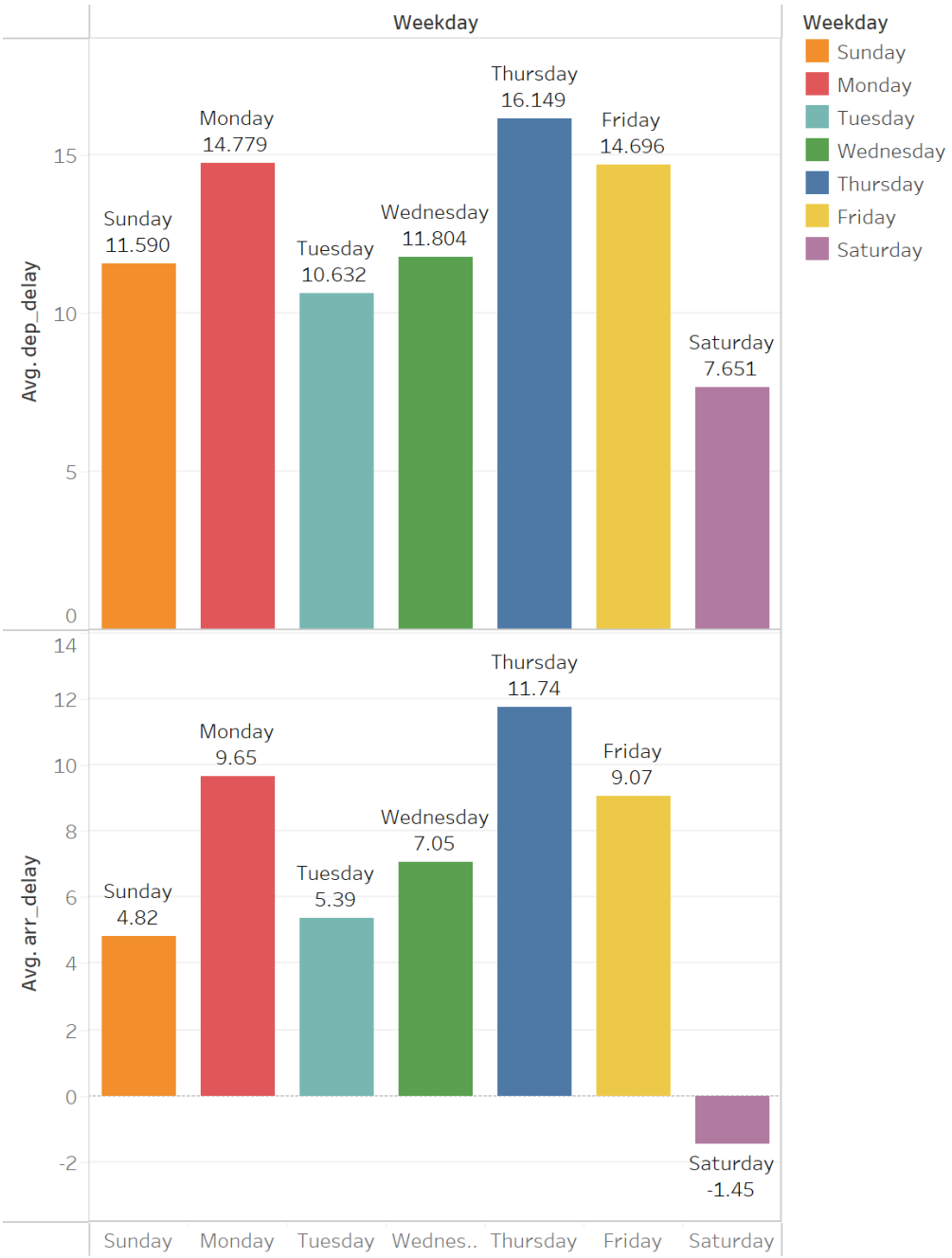
F9 has average 20.22 minutes departure delay and 21.92 minutes arrival delay, which is in the worst position among other airlines.

US has 3.78 and 2.13 minutes respectively, still need to improve, but so far it has the shortest delay time compare with other carriers.

3. Avoid Thursdays and Any Rush Hours.

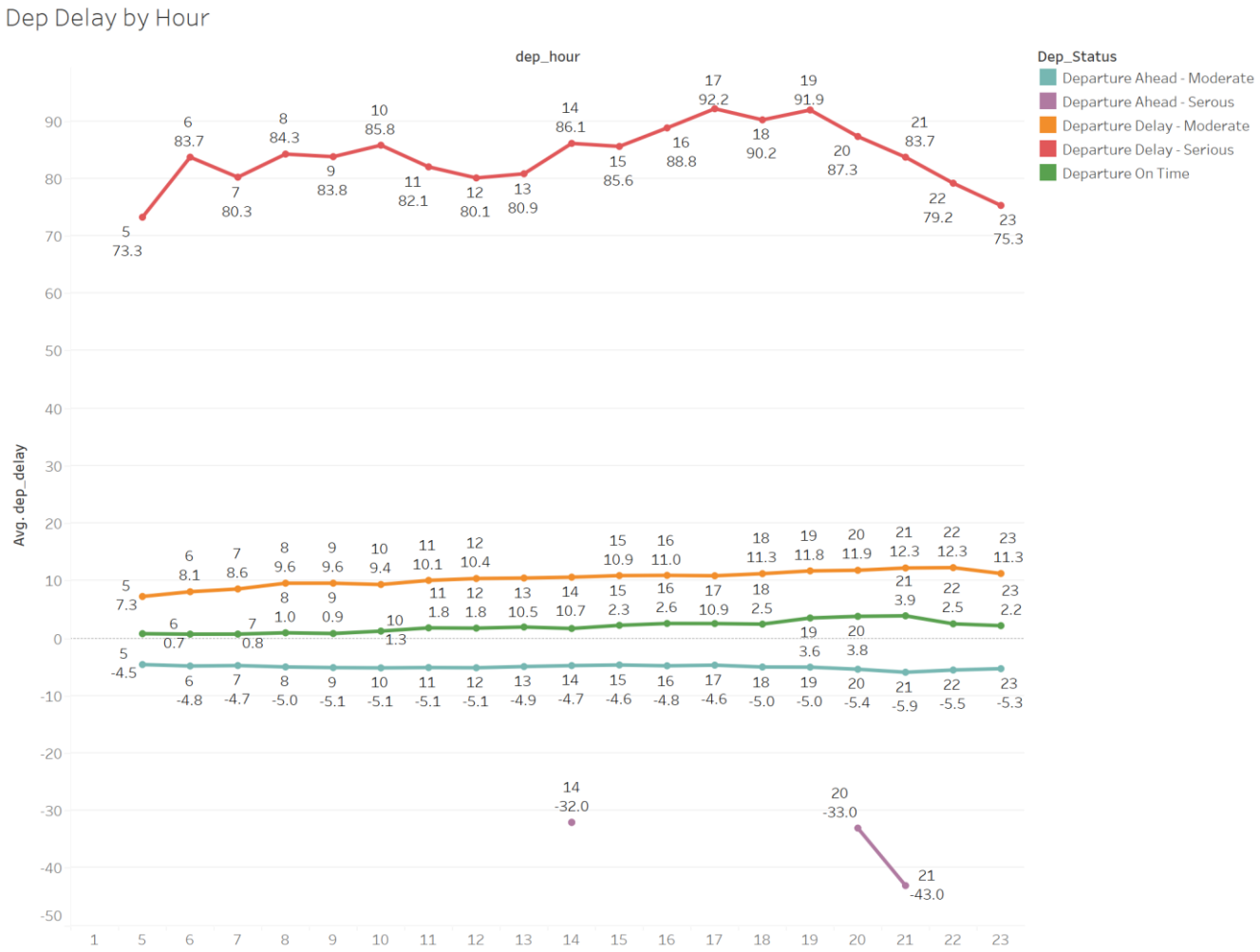
Avoid Thursdays as Saturdays are calmer.

Dep Delay by Weekday



Average of dep_delay and average of arr_delay for each Weekday. Color shows details about Weekday. The marks are labeled by Weekday. For pane Average of dep_delay: The marks are labeled by Weekday and average of dep_delay. For pane Average of arr_delay: The marks are labeled by Weekday and average of arr_delay. The data is filtered on Dep_Status, time_hour Weekday and origin. The Dep_Status filter keeps Departure Ahead - Moderate, Departure Ahead - Serous, Departure Delay - Moderate, Departure Delay - Serious and Departure On Time. The time_hour Weekday filter keeps 7 of 7 members. The origin filter keeps EWR, JFK and LGA.

When it is rush hours on the road, it is same at the airport.



The trend of average of dep_delay for dep_hour. Color shows details about Dep_Status. The marks are labeled by dep_hour and average of dep_delay. The data is filtered on time_hour Weekday and origin. The time_hour Weekday filter keeps 7 of 7 members. The origin filter keeps EWR, JFK and LGA. The view is filtered on Dep_Status, which keeps Departure Ahead - Moderate, Departure Ahead - Serous, Departure Delay - Moderate, Departure Delay - Serious and Departure On Time.

(The End)

Reference:

1. https://www.claimcompass.eu/blog/why-is-my-flight-delayed/

2. https://www.claimcompass.eu/blog/airline-rating-2019-insights/

Sales Performance Data Cleanup

USING TABLEAU PREP

BY PENG WANG

Metro College of Technology
Instructor: Mr. Hamid Rajaee

METRO COLLEGE OF TECHNOLOGY

Business Scenarios

- This is the performance data for sales team (500 sales reps) collected between Jan 2014 to December 2018.
- Scenario: potential customers have rather spontaneous demand. When this happens, sales team puts an order lead in the system.
- Sales reps then try to get set up a meeting that occurs around the time the order lead was noticed. Sometimes before, sometimes after.
- Sales reps have an expense budget and always combine the meeting with a meal for which they pay.

Project Scope

To analysis, I have three data sources:

1. order leads (contains all the order leads and conversion info) :
    total observations: 100,000
    Key Attributes: order date, order value, converted (0/1)
2. sales team (includes companies and responsible sales
    total observations: 4619
    Key Attributes: Sale Rep, Company Name
3. invoices (provides information on the invoices and participants)
    total observations: 49990
    Key Attributes: Date/time of Meal , Meal Price

Descriptive Analysis

- Overall Conversion Rate development
- Conversion Rate by Sales Reps by years

Exploratory Data Analysis

- Effect of meals
- Impact of type of meal

- Impact of timing of meal before or after the order lead

Conclusion/Recommendation

- to increase sales team's performance

Data Manipulating

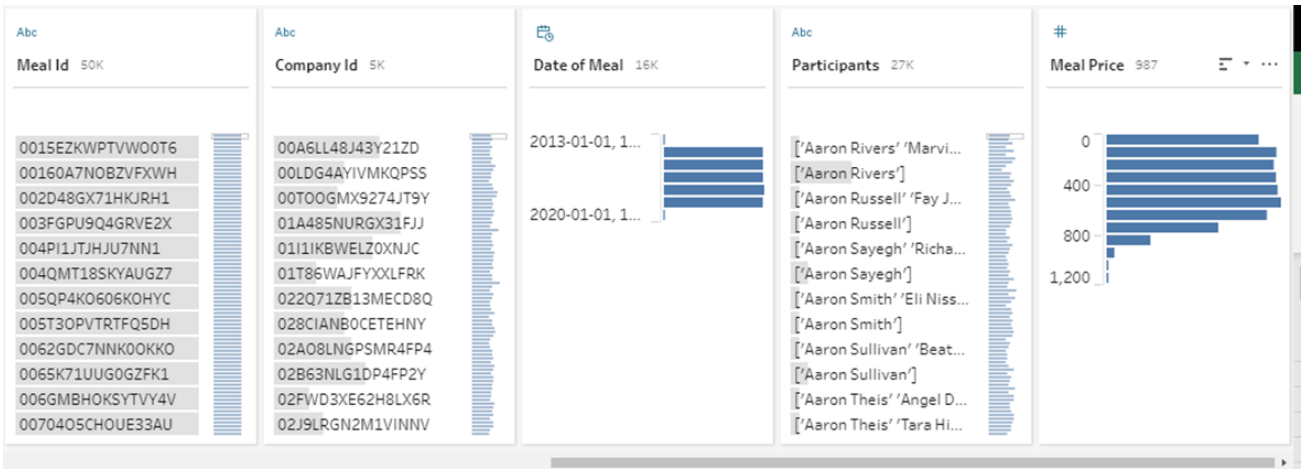
To analysis the impact of meals related with order leads conversion rate, I need to categorize each meal in to certain type, based on the time, date, and participant of the meal.

What 'meal type' I need?

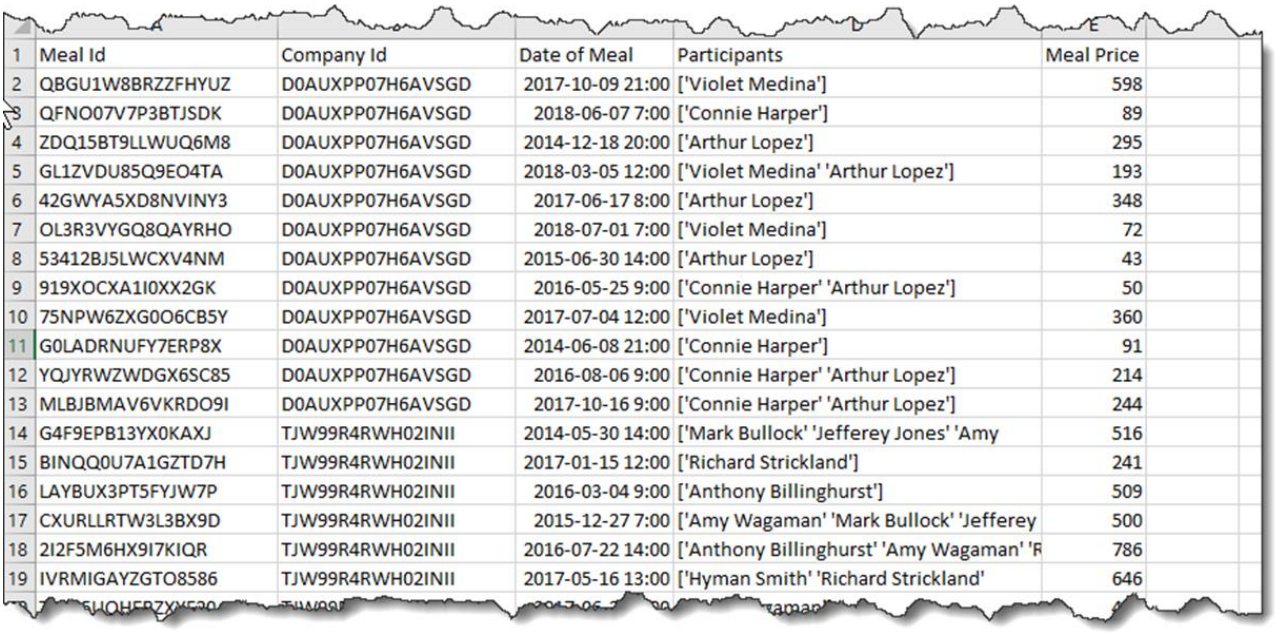
- Individual or Group;
- $ spent per person;
- Brunch, Lunch, or Dinner;
- Weekday, Weekend, or Holiday;

Challenges

1. 'invoices.csv' does not have meal type.



2. The meal participants' names were mixed into one column.

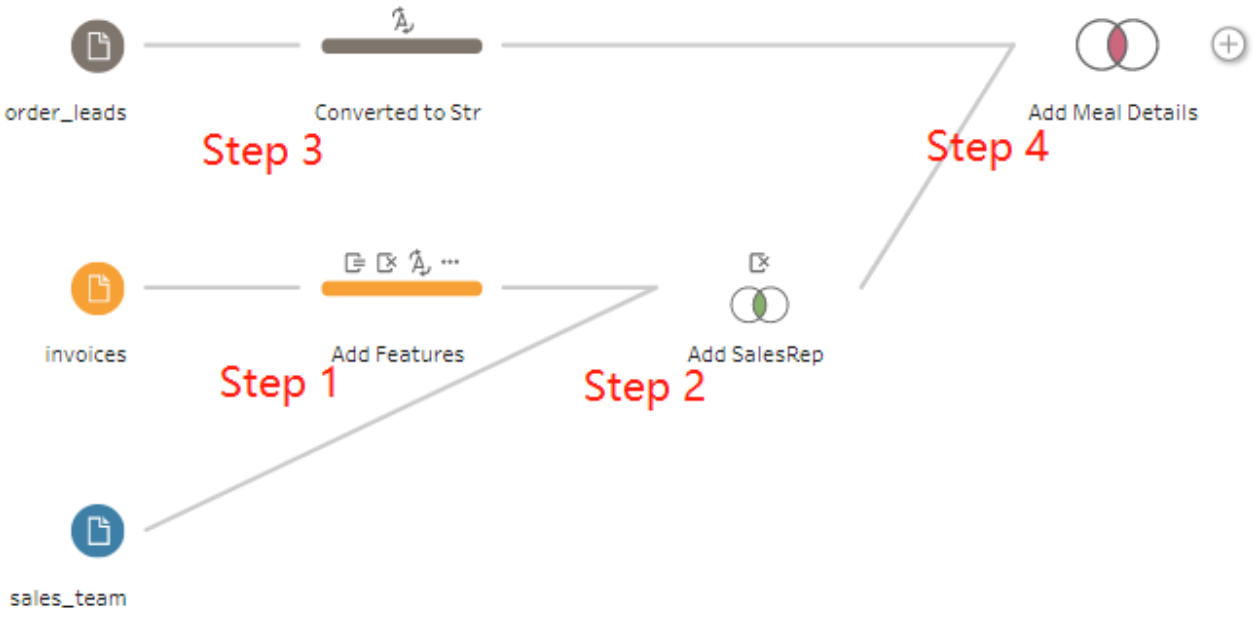


	A	B	C	D	E
1	Meal Id	Company Id	Date of Meal	Participants	Meal Price
2	QBGU1W8BRZZFHYUZ	D0AUXPP07H6AVSGD	2017-10-09 21:00	['Violet Medina']	598
3	QFNO07V7P3BTJSDK	D0AUXPP07H6AVSGD	2018-06-07 7:00	['Connie Harper']	89
4	ZDQ15BT9LLWUQ6M8	D0AUXPP07H6AVSGD	2014-12-18 20:00	['Arthur Lopez']	295
5	GL1ZVDU85Q9EO4TA	D0AUXPP07H6AVSGD	2018-03-05 12:00	['Violet Medina' 'Arthur Lopez']	193
6	42GWYA5XD8NVINY3	D0AUXPP07H6AVSGD	2017-06-17 8:00	['Arthur Lopez']	348
7	OL3R3VYGQ8QAYRHO	D0AUXPP07H6AVSGD	2018-07-01 7:00	['Violet Medina']	72
8	53412BJ5LWCXV4NM	D0AUXPP07H6AVSGD	2015-06-30 14:00	['Arthur Lopez']	43
9	919XOCXA1I0XX2GK	D0AUXPP07H6AVSGD	2016-05-25 9:00	['Connie Harper' 'Arthur Lopez']	50
10	75NPW6ZXG0O6CB5Y	D0AUXPP07H6AVSGD	2017-07-04 12:00	['Violet Medina']	360
11	G0LADRNUFY7ERP8X	D0AUXPP07H6AVSGD	2014-06-08 21:00	['Connie Harper']	91
12	YQJYRWZWDGX6SC85	D0AUXPP07H6AVSGD	2016-08-06 9:00	['Connie Harper' 'Arthur Lopez']	214
13	MLBJBMAV6VKRDO9I	D0AUXPP07H6AVSGD	2017-10-16 9:00	['Connie Harper' 'Arthur Lopez']	244
14	G4F9EPB13YX0KAXJ	TJW99R4RWH02INII	2014-05-30 14:00	['Mark Bullock' 'Jefferey Jones' 'Amy	516
15	BINQQ0U7A1GZTD7H	TJW99R4RWH02INII	2017-01-15 12:00	['Richard Strickland']	241
16	LAYBUX3PT5FYJW7P	TJW99R4RWH02INII	2016-03-04 9:00	['Anthony Billinghurst']	509
17	CXURLLRTW3L3BX9D	TJW99R4RWH02INII	2015-12-27 7:00	['Amy Wagaman' 'Mark Bullock' 'Jefferey	500
18	2I2F5M6HX9I7KIQR	TJW99R4RWH02INII	2016-07-22 14:00	['Anthony Billinghurst' 'Amy Wagaman' 'R	786
19	IVRMIGAYZGTO8586	TJW99R4RWH02INII	2017-05-16 13:00	['Hyman Smith' 'Richard Strickland'	646

Solutions

Add new columns to hold below features:

- Split ‘Participants’
- Aggregate ‘Meal Price’
- Split ‘Date of Meal’
- Match meal type and weekday
- Match holiday

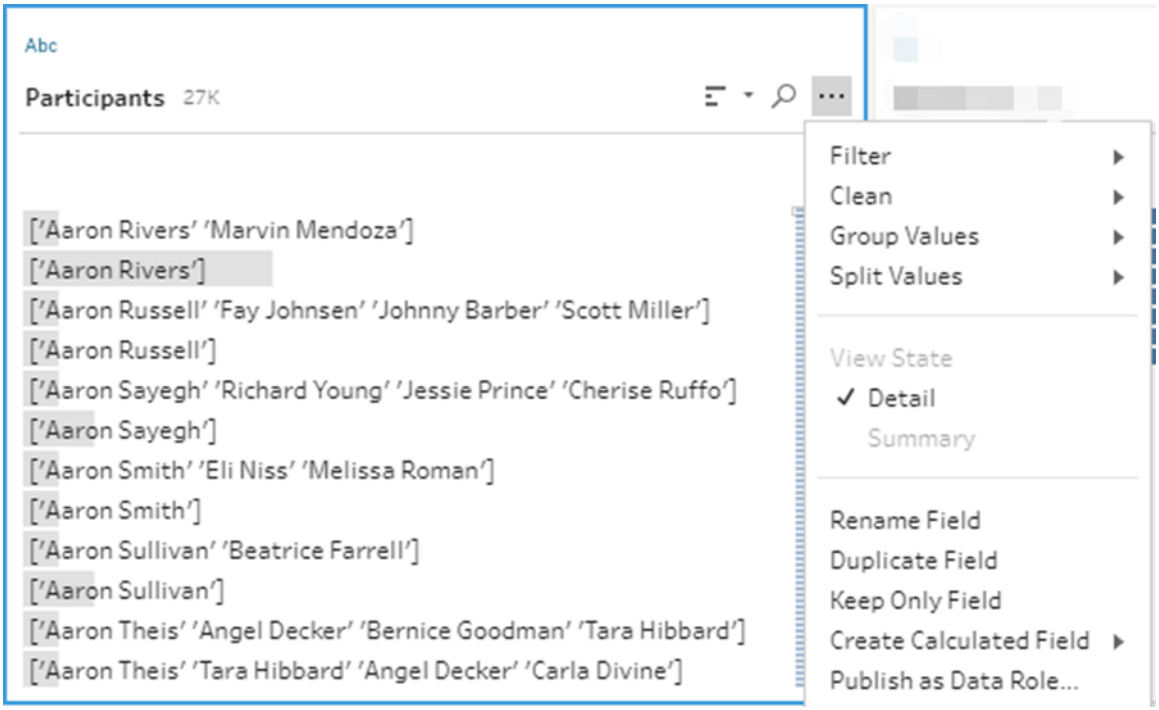


Split 'Participants'

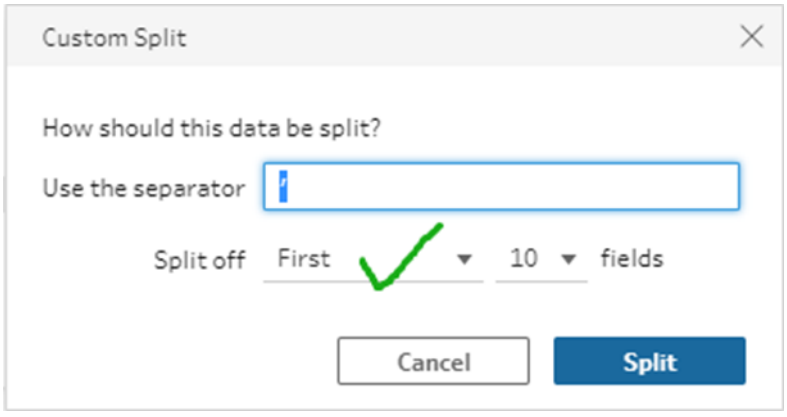
In order to count the number of participants, I need splitting them from multiple names in one column, to one name per column.

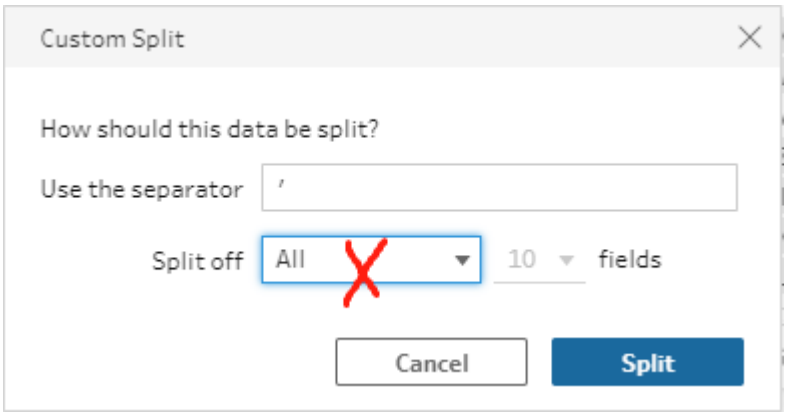
Step 1: Clean > Remove Letters
To see how many delimiters involved.

Step 2: Split Values > Customer Split



Ensure to select 'First' from Split off drop down list:

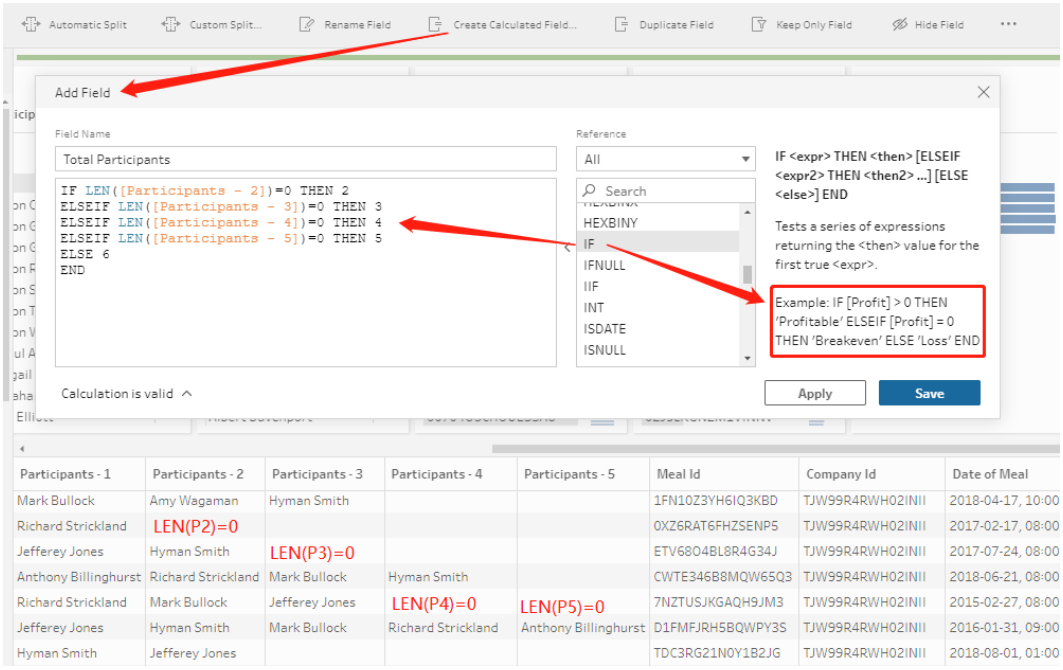




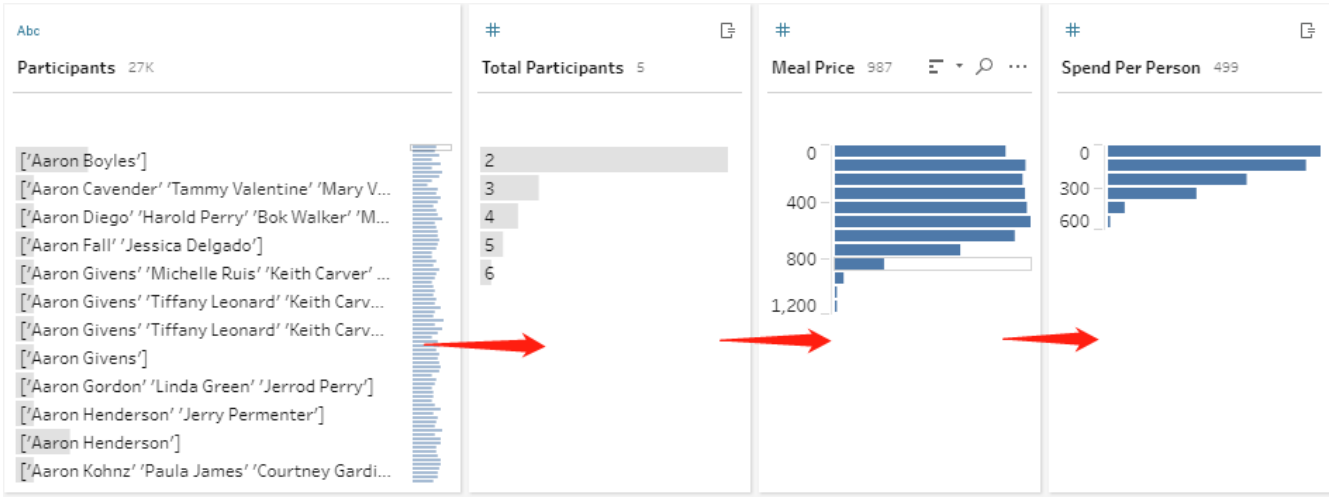
Aggregate 'Meal Price'

Once participants names split, it's the time to count the participants, then calculate the meal cost per participant.

Step 1: Create a Calculated Field to count the number of meal participants (include sales rep).

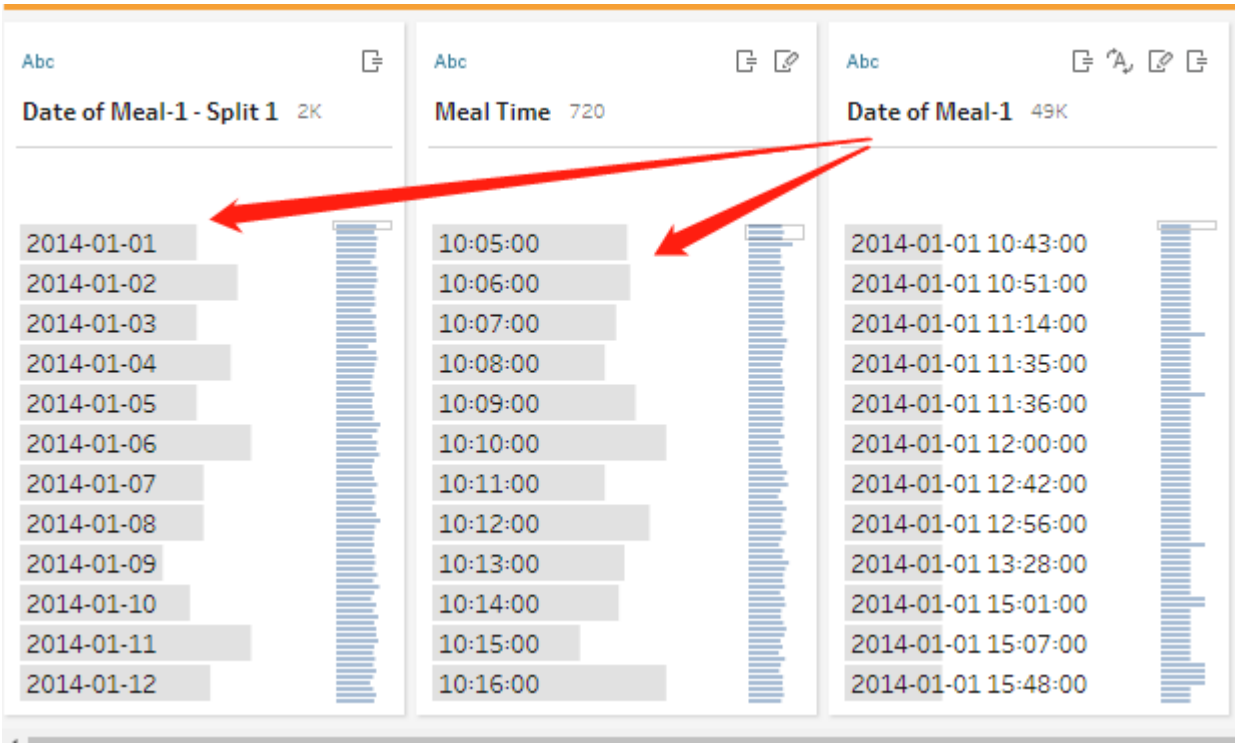


Step2: Add another Calculated Field to get the Spend Per Person.



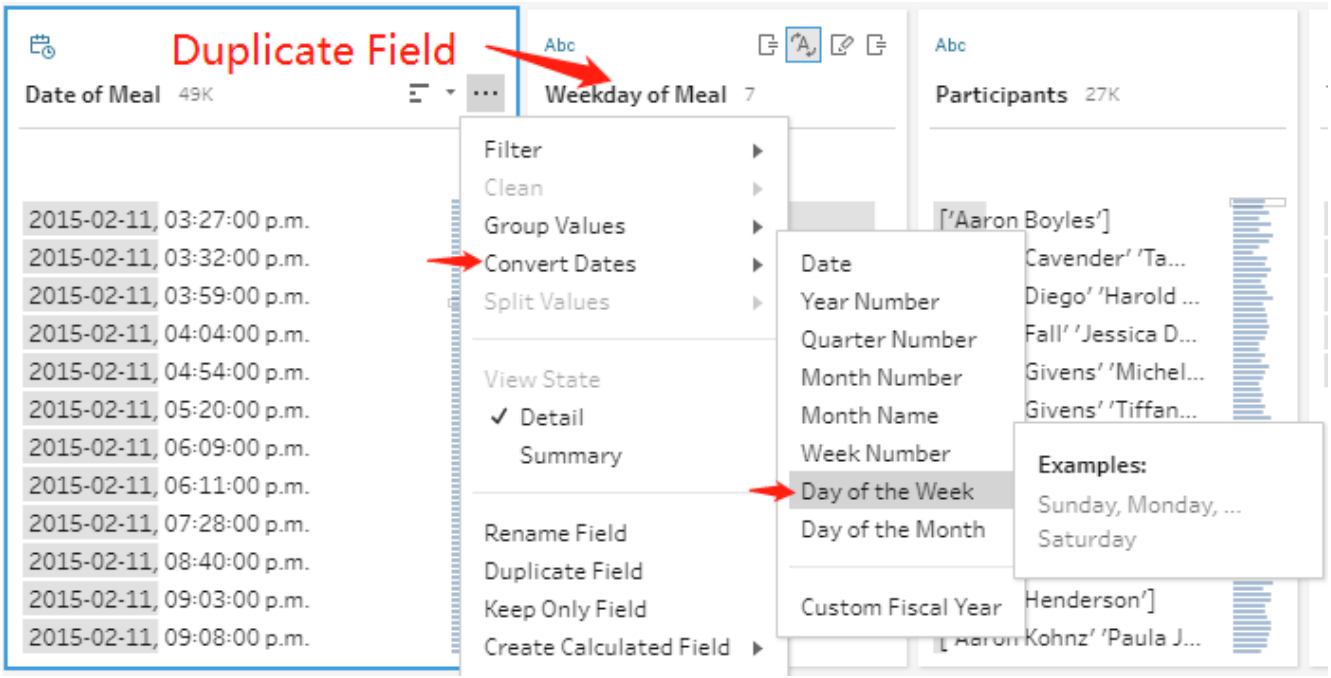
Split 'Date of Meal'

Split Date of Meal in order to find Weekday, and according to the meal time, define Brunch, Lunch or Dinner.

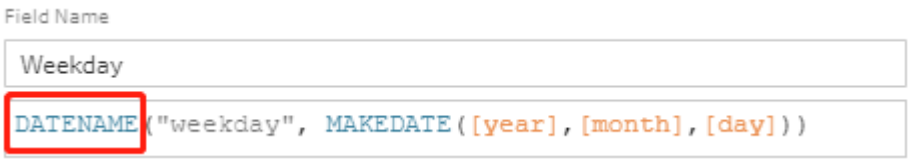


There are two methods to extract weekday out of the source data.

Method 1



Method 2



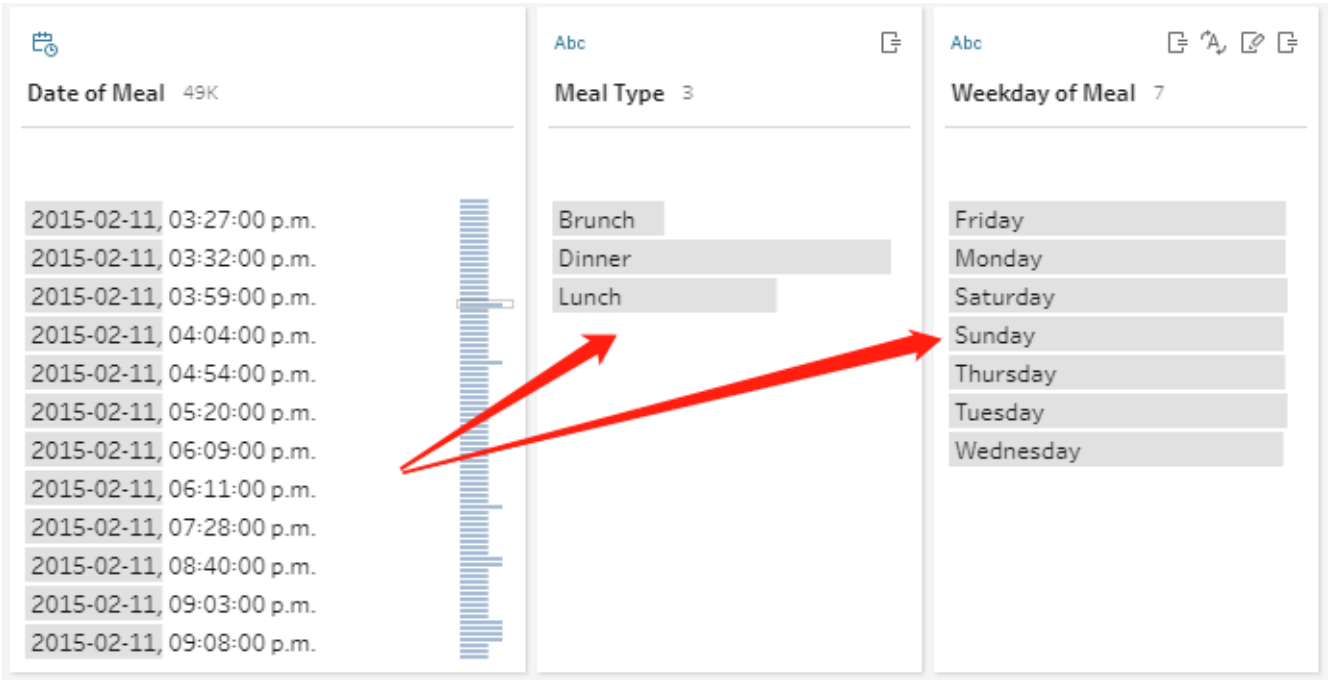
Match meal type and weekday

To define the meal type, any meal time in the morning will be Brunch; in the pm will be Lunch, and Dinner otherwise.

Field Name

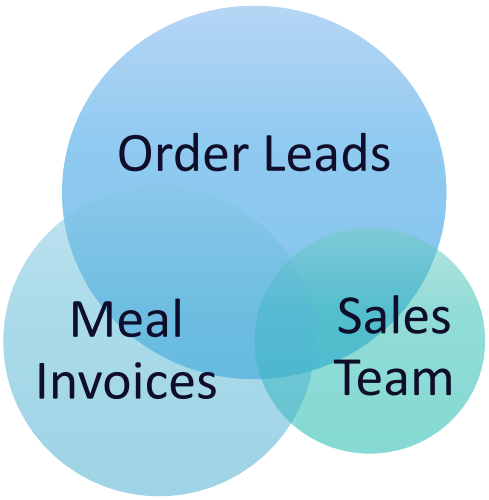
Meal Type

```
IF
STARTSWITH([Meal Time],"10")OR
STARTSWITH([Meal Time],"11")
THEN "Brunch"
ELSEIF
STARTSWITH([Meal Time],"12")OR
STARTSWITH([Meal Time],"13")OR
STARTSWITH([Meal Time],"14")OR
STARTSWITH([Meal Time],"15")
THEN "Lunch"
ELSE "Dinner"
END
```



Tables Joining

Three tables can be connected by linking the primary keys.



order_leads 7 fields | Filter Values...

Clear the check box to remove fields. You can

Fields selected: 7 of 7

	Type	Field Name
☑	Abc	Order Id
☑	Abc	Company Id
☑	Abc	Company Name
☑	📅	Date
☑	#	Order Value
☑	#	Converted
☑	Abc	Meal Id

invoices 5 fields | Filter Values...

Clear the check box to remove fields. You can

Fields selected: 5 of 5

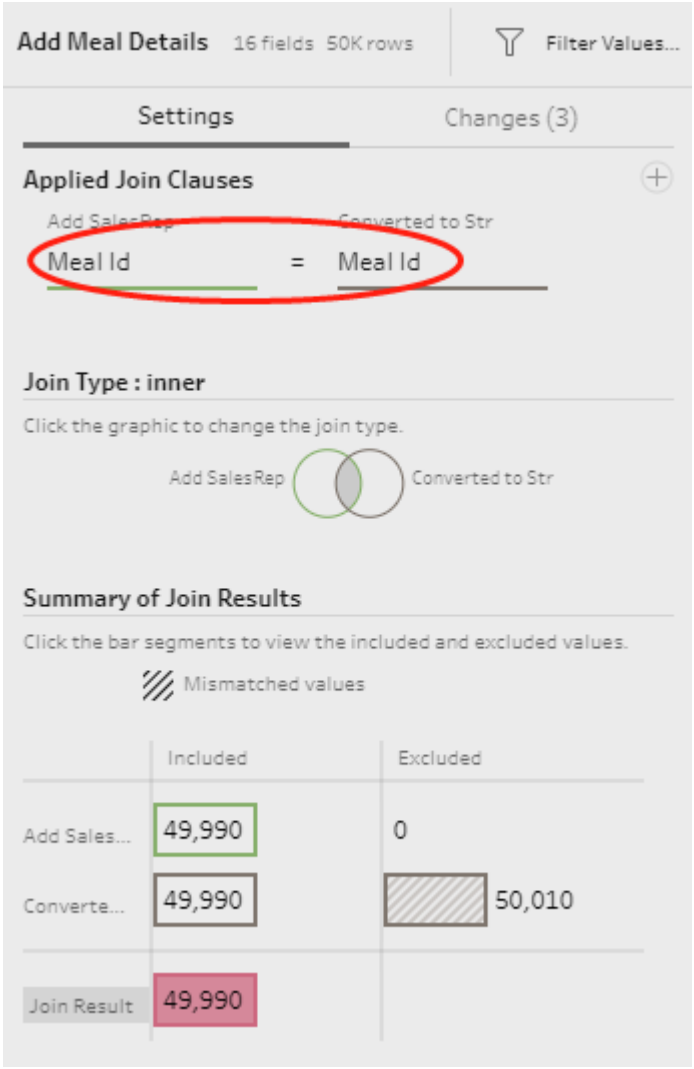
	Type	Field Name
☑	Abc	Meal Id
☑	Abc	Company Id
☑	📅	Date of Meal

sales_team 4 fields | Filter Values...

Clear the check box to remove fields. You can

Fields selected: 4 of 4

	Type	Field Name
☑	Abc	Sales Rep
☑	Abc	Sales Rep Id



Below is the final version of consolidated dataset ready for further analysis.

Sales Rep	Company Name	Order Date	Order Value	Converted	Meal Id	Date of Meal	Meal Type	Weekday of Meal	Participants	Total Participants	Meal Price	Spend Per Person
Angelo Travis	Pursuant Spurdle	2017-04-08	4,108	1	XIR5XP8BL	2017-04-17, 12:35:00 p.m.	Lunch	Monday	['Jennie Starks']	2	507	254
Enrique Burns	Littlest Neighbors	2017-04-08	1,191	0	A8Z8Y6EL8	2017-04-17, 12:59:00 p.m.	Lunch	Monday	['Michael Scott']	2	60	30
Dominic Williar	Good-Looking Poly	2017-04-08	2,158	0	CYKO4CYSC	2017-04-17, 01:01:00 p.m.	Lunch	Monday	['Marie Belton']	2	42	21
Kathy Webb	Secretarial Mattre	2017-04-08	8,141	0	MBCA4F0R	2017-04-17, 01:53:00 p.m.	Lunch	Monday	['Willie Prout']	2	421	211
Paul Solomon	Unqualified Geolo	2017-04-09	7,462	0	CDJLR0AEN	2017-04-17, 03:13:00 p.m.	Lunch	Monday	['Peter Willis']	2	661	331
Sharon Gellings	Interlibrary Lassw	2017-04-09	5,409	0	MXKG8537	2017-04-17, 05:24:00 p.m.	Dinner	Monday	['Richard Padgett']	2	770	385
Lynda Harvey	Guatemalan Musc	2017-04-09	6,721	0	HR6PWPUI	2017-04-17, 06:22:00 p.m.	Dinner	Monday	['Christina Baker']	2	388	194
John Piazza	Unstructured Illus	2017-04-09	4,579	1	X1K1S0K0S	2017-04-17, 06:25:00 p.m.	Dinner	Monday	['Eva Hasty' 'Lewis Coleman']	3	525	175
John Piazza	Airless Vice-Chairr	2017-04-09	6,643	1	HD9TO5ML	2017-04-17, 06:53:00 p.m.	Dinner	Monday	['Howard Hiltner']	2	581	291
Christopher Vil	Furtive Insurance	2017-04-09	6,637	0	85EMUCVH	2017-04-17, 07:00:00 p.m.	Dinner	Monday	['Roy Ross']	2	441	221
Mark Brown	Bestselling Holes	2017-04-09	2,606	0	ZVX4VLDVC	2017-04-17, 07:37:00 p.m.	Dinner	Monday	['Tony Bertalan' 'Dennis Jaran	5	90	18

(The End)