



Market Match

FINAL PROJECT

NAYA COLLAGE

REVITAL MATITYAHU & ERAN SARID

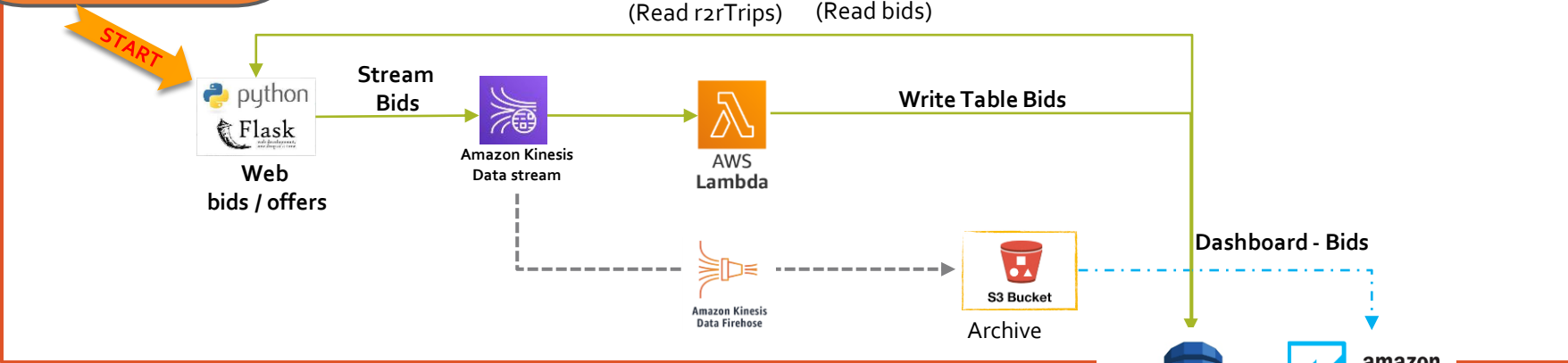
08/2021

Introduction

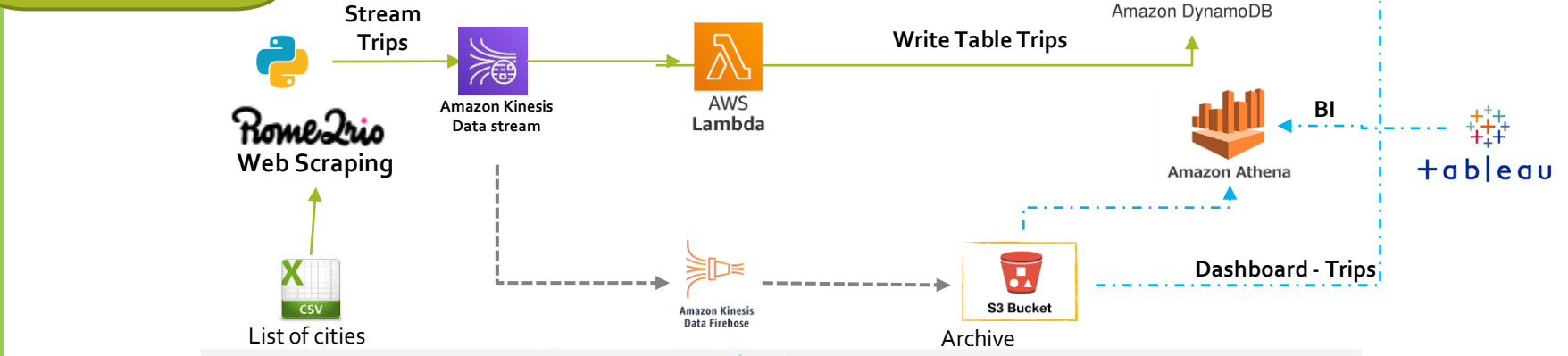
- A web application for online commercial between consumers and consumers/suppliers.
- Consumers can add multiple bids in several categories.
- Consumers/Suppliers can make offers
- Get enrichment from related web sites

Market Match Architecture

Main data pipeline



Info - Enrichment



AiFlow

- We use EC2 to runs AirFlow web service and scheduler
- It executes multiple DAGS that bring up “Market-Match” application and Web Scrapping processes

All 36			Active 0	Paused 36
i	DAG	Owner		
<input type="checkbox"/>	dag_load_bids1	tyche		
<input type="checkbox"/>	dag_load_market	tyche		
<input type="checkbox"/>	dag_load_r2rtrips	tyche		



S3

- Store json files for archive and analytics

Amazon S3

▼ **Account snapshot**
Last updated: Aug 20, 2021 by Storage Lens. Metrics are generated every 24 hours. [Learn more](#)

Total storage	Object count
198.6 MB	1.7 k

Buckets (18) [Info](#)
Buckets are containers for data stored in S3. [Learn more](#)

Q BID

	Name ▲	AWS Region ▼
<input type="radio"/>	bidy-bids	US East (Ohio) us-east-2
<input type="radio"/>	bidy-scripts	US East (Ohio) us-east-2



Amazon S3 > bidy-bids > bids2021/ > 08/ > 17/ > 12/

12/

Objects | Properties

Objects (1)
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your I

Q Find objects by prefix

<input type="checkbox"/>	Name
<input type="checkbox"/>	firehose_bids_kinesis_s3-1-2021-08-17-12-25-46-0de6bcd3-1e0e-4991-a0ac-c7287bd12b5d

Cloud9

[AWS Cloud9](#) > [Environments](#) > ec2cloud9

ec2cloud9

[Open IDE](#) [Edit](#) [Delete](#)

Environment details

Name	EC2 instance type	Security groups	Environment ARN
ec2cloud9	t2.micro	sg-026f2edab00a63609	arn:aws:cloud9:us-east-2:256760239101:environment:9610d25de76b427495dbf343903846ab
Description	Memory	VPC	Number of members
ide for DynamoDB	1 GiB	vpc-fb6d0e90	2
Type	vCPU	Subnet	Lifecycle State
EC2	1	subnet-f895bcb4	CREATED
Owner ARN	Storage	EC2 Instance	
arn:aws:iam::256760239101:root	EBS only	Go To Instance	
		Environment path	
		<code>/home/ec2-user/environment</code>	

Tags (0)



DynamoDB

- bids table: Users \ Bids \ Offers

bids

user_name (Partition key)

Guglielmo530MartagonGEN

entity_id (Sort key)

Equal to ☐ Sort descending

Filters

Run Reset

Completed Read capacity units consumed: 0.5

Items returned (9)

	user_name	entity_id	bid	bid_info	category	contact
<input type="checkbox"/>	Guglielmo530MartagonGEN	BID#01FDERF3NPCMQ5E...		{ "target_pr...	flights	
<input type="checkbox"/>	Guglielmo530MartagonGEN	BID#01FDERF3QAM125CZ...		{ "target_pr...	flights	
<input type="checkbox"/>	Guglielmo530MartagonGEN	CUST#Guglielmo530Mart...				
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#01FDERF7412KJC...	BID#01FDERF3NPCM...		{ "nam	
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#01FDERF7DJ74EF...	BID#01FDERF3QAM12...		{ "nam	
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#01FDFCXHSEYERT...	BID#01FDERF3NPCM...		{ "nam	
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#01FDFCXHH9E73...	BID#01FDERF3QAM12...		{ "nam	
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#Kayak#01FDFDHX...	BID#01FDERF3QAM12...		{ "nam	
<input type="checkbox"/>	Guglielmo530MartagonGEN	OFFER#TripAdvisorFlights...	BID#01FDERF3NPCM...		{ "nam	

Item editor

Attributes

Attribute name	Value	Type	
user_name - Partition key	Guglielmo530MartagonGEN	String	
entity_id - Sort key	CUST#Guglielmo530MartagonGEN	String	
password	1	String	Remove
fname	Guglielmo	String	Remove
timestamp	2021-08-19T08:29:25.938832	String	Remove
lname	Martagon	String	Remove
email	Guglielmo_Martagon@gmail.com	String	Remove
phone	193-167-9912	String	Remove
create_date	2020-08-07T12:50:25.860809	String	Remove
update_date	2020-08-07T12:50:25.860809	String	Remove

Add new attribute



DynamoDB

- r2rTrips table: trips options-transits enrichment
- r2rTrips table: hierarchy arrays analysis

Time to Live (TTL) [Info](#)

Automatically delete expired items from a table.

[Run preview](#)[Disable](#)

TTL status

✓ Enabled

TTL attribute

timestamp [View items](#) [View graph](#)

Items deleted in the last 24 hours

0 [View graph](#)

trip	entity_id	arrays_...	info	timestamp (TTL)	total_arr...	total_em...	total_entries
Sao Paulo	STATS#Beijing	24	{ "total_arr...	2021-08-28T09:15:14.9...	784	7021	1166
Sao Paulo	STATS#Cairo	24	{ "total_arr...	2021-08-28T09:15:14.9...	778	6105	1172
Sao Paulo	STATS#Delhi	24	{ "total_arr...	2021-08-28T09:15:14.9...	772	5460	1178

trip	entity_id	info_transits	timestamp (TTL)
Sao Paulo	DEST#Delhi#OPTION0	{ "org" : { "S" : "São Paulo" }, "dest" : { "S" : "Delhi" }, "dest_...	2021-08-28T21:11:46.777619
Sao Paulo	DEST#Delhi#OPTION1	{ "org" : { "S" : "São Paulo" }, "dest" : { "S" : "Delhi" }, "dest_...	2021-08-28T21:11:46.777619
Sao Paulo	DEST#Jakarta#OPTION0	{ "org" : { "S" : "São Paulo" }, "dest" : { "S" : "Jakarta" }, "des...	2021-08-28T21:11:46.777619



Kinesis Data Streams

- Pipeline for transfer data through Lambda to DynamoDB
- Bids app data
- Rome2Rio scrapping data

Amazon Kinesis > Data streams

Data streams (3) [Info](#) Processes ▾ Actions ▾ Create data stream

<input type="checkbox"/>	Data stream name ▾	Status ▾	Open shards ▾	Data retention period ▾	Encryption ▾	Consumers with enhanced fan-out ▾
<input type="checkbox"/>	stream_r2rTrips_kinesis_dynamodb	✔ Active	1	1 day	Disabled	0
<input type="checkbox"/>	stream_bids_kinesis_dynamodb	✔ Active	1	1 day	Disabled	0



Lambda

- Write bids and trips Items into DynamoDB tables

Lambda > Functions

Functions (3)

Filter by tags and attributes or search by keyword 2 matches


"DYNA" X Clear filters


	Function name	Description	Package type	Runtime
<input type="radio"/>	bids_kinesis_dynamodb_lambda	-	Zip	Python 3.8
<input type="radio"/>	r2rTrips_kinesis_dynamodb_lambda	-	Zip	Python 3.8


Lambda > Functions > bids_kinesis_dynamodb_lambda

bids_kinesis_dynamodb_lambda

▼ Function overview Info

 bids_kinesis_dynamodb_lambda

 Layers (0)

 Kinesis

+ Add trigger

+ Add destination

Code Test Monitor Configuration Aliases Versions

Code source Info

File Edit Find View Go Tools Window Test Deploy Changes deployed

Go to Anything (Ctrl-P)

Environment

- bids_kinesis_dynam...
- lambda_function.py

```
1 import datetime
2
3
4
5
6
7
8 def lambda_handler(event, context):
9     """
10     Receive a batch of events from Kinesis and insert into our DynamoDB table
11     """
12     print('Received request')
13     item = None
14     dynamo_db = boto3.resource('dynamodb')
15     table = dynamo_db.Table('bids')
16     decoded_record_data = [base64.b64decode(record['kinesis']['data']) for record in event['Records']]
17     deserialized_data = [json.loads(decoded_record) for decoded_record in decoded_record_data]
18
19     with table.batch_writer() as batch_writer:
20         for item in deserialized_data:
21             print(item)
22             # Add a processed time so we have a rough idea how far behind we are
23             batch_writer.put_item(Item=item)
24
25     # Print the last item to make it easy to see how we're doing
26     print(json.dumps(item))
27     print('Number of records: {}'.format(len(deserialized_data)))
```



Kinesis Firehose

- Store data streams events to S3 for archive and analytics

Amazon Kinesis > Delivery streams

Delivery streams (2) 🔄 Delete Create delivery stream

	Name ▲	Status ▼	Creation time ▼	Source ▼	Data transformat... ▼	Destination ▼	Destinatio... ▼
<input type="radio"/>	firehose_bids_kinesi...	Active	August 17, 2021, 15...	stream_bids_kinesis...	Not enabled	Amazon S3	body-bids
<input type="radio"/>	firehose_r2rTrips_ki...	Active	August 11, 2021, 02...	stream_r2rTrips_kin...	Not enabled	Amazon S3	r2rtrips-archi...

Name ▲	Type ▼	Last modified ▼
firehose_r2rTrips_kinesis_s3-1-2021-08-24-21-30-43-eb4bbfd9-f9d5-4d24-a423-f5f046919a1d	-	August 25, 2021, 00:35:45 (UTC+03:00)
firehose_r2rTrips_kinesis_s3-1-2021-08-24-21-39-25-64b0ff0c-bccc-4f77-9c5e-420793b7bdda	-	August 25, 2021, 00:44:28 (UTC+03:00)
firehose_r2rTrips_kinesis_s3-1-2021-08-24-21-45-34-ef91487e-88ac-42e4-92d3-a6a9757f8894	-	August 25, 2021, 00:50:36 (UTC+03:00)
firehose_r2rTrips_kinesis_s3-1-2021-08-24-21-52-17-8943f861-ed51-4b2a-8b80-7327e7847833	-	August 25, 2021, 00:57:19 (UTC+03:00)
firehose_r2rTrips_kinesis_s3-1-2021-08-24-21-59-12-3e419219-9a06-423d-82e3-2bf6206bf9bc	-	August 25, 2021, 01:04:15 (UTC+03:00)



Python - Flask

- “Market-Match” application

Market Match Place

Start purchasing by clicking the link below



Revital Matityahu & Eran Sarid
version - 2

Login

Please Create your Account

User Name

First Name

Last Name

Phone Number

City

Address

Email Address

Password

Confirm Password

Already have an account?

Login

Create Account

Flask Market Home Market

Welcome Logout Login Register

Success! You are logged in as: Kyle825LobilloGEN

Please Create your Bids

Flight Select Origin... Select Destination...

adults children mm/dd/yyyy mm/dd/yyyy

target_price

Create Request

My Bids

category	origin	destination	From Date	To Date	adults	children	target_price		
Flights	Tel Aviv	New York	15.9	30.9	2	3	700	More Info	More Info
Flights	Tel Aviv	Bangkok	15.9	30.9	2	3	735	More Info	More Info
Flights	Tel Aviv	London	15.9	30.9	2	3	350	More Info	More Info
Flights	Tel Aviv	Moscow	15.9	30.9	2	3	310	More Info	More Info

Make Offer To Other Bids



Python - Web Scrapping

- Scrapping rome2rio site for enrichment trips options

Tel Aviv, Israel → London, England

Partially open COVID-19 Guide: International entry to United Kingdom is allowed for permanent residents and citizens, as well as. [Show me more](#)

RECOMMENDED FASTEST

Fly to London Heathrow
7h 54m From US\$100

SECOND FASTEST

Fly to London Gatwick, train
8h 29m From US\$116

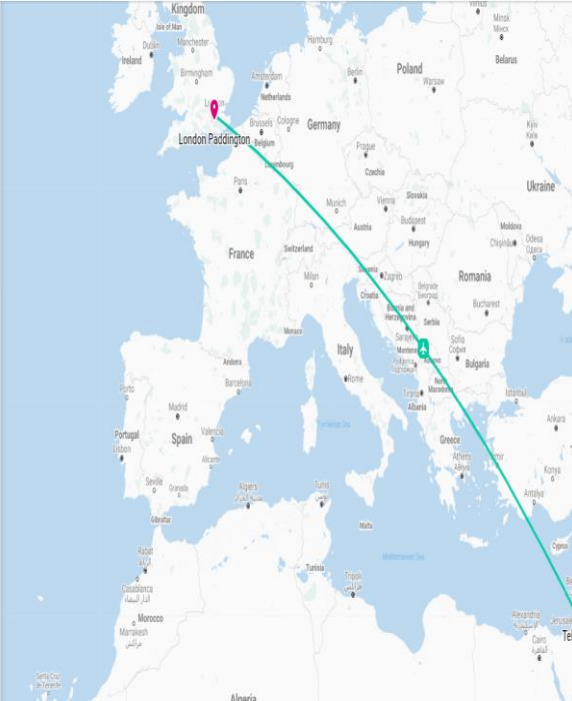
Fly to London City
10h 6m From US\$166

CHEAPEST

Fly to London Luton, train
8h 41m From US\$67

SECOND CHEAPEST

Fly to London Stansted
10h 8m From US\$80



Extracted original Json from site

```
1
2 Tel Aviv to London - 6 ways to travel via train, and plane
3 False
4
5 1 0
6   1 node
7   2 Tel Aviv
8   3 32.08088
9   4 34.78057
10  5 IL
11  6 Asia/Jerusalem
12  7 2.5
13
14 1 node
15 2 London Paddington
16 3 51.51574
17 4 -0.17554
18 5
19 6
20 7 -1
21
22 4 Fly-to-London-Heathrow
23 5 Fly to London Heathrow
24 6 28427.708
25 7 20940
26 8 7487.708
27
28 1 transit
29 2 train
30 3 purple
31 4 1140
32 5 871.47
33 6 14.43004
34
35 1 station
36 2 Tel Aviv Center
37 3 32.08371
38 4 34.78057
```

• Python - Web Scrapping

Extracted Flight info

```
{
  "transits": [
    {
      "transit": {
        "type": "purple",
        "min_price_unit": 2,
        "dest_cat": "station",
        "org_cat": "station",
        "duration": 1140,
        "org_lat": "32.08371",
        "price_unit": "GBP",
        "dest_lat": "32.00036",
        "cat": "transit",
        "dest_name": "BG-Airport",
        "dist_option_dest_to_transit_dest": 3577,
        "max_price_unit": "3.59",
        "name": "train",
        "dist_option_orig_to_transit_orig": 2,
        "org_name": "Tel Aviv Center",
        "dist_transit": 11,
        "org_lng": "34.79825",
        "dest_lng": "34.87048"
      }
    },
    {
      "transit": {
        "type": "grey-walk",
        "dest_cat": "airport",
        "org_cat": "station",
        "duration": 0,
        "org_lat": "32.00036",
        "dest_lat": "32.00025",
        "cat": "walk",
        "dest_name": "Tel Aviv (TLV)",
        "dist_option_dest_to_transit_dest": 3577,
        "name": "transfer"
      }
    }
  ]
}
```

transits

[-] transits [6]

[-] transit {}

[+] transit {}

[+] transit {}

[-] transit {}

type	turquoise
dest_zone	EU
min_price_unit	65
dest_town	England
unknown	
org_lat	32.00025
org_town	HaMerkaz
dest_code	LHR
price_unit	GBP
dest_lat	51.47034
org_country	Israel
cat	flight
dest_name	London Heathrow
dist_option_dest_to_transit_dest	20
max_price_unit	264.82
org_zone	AS

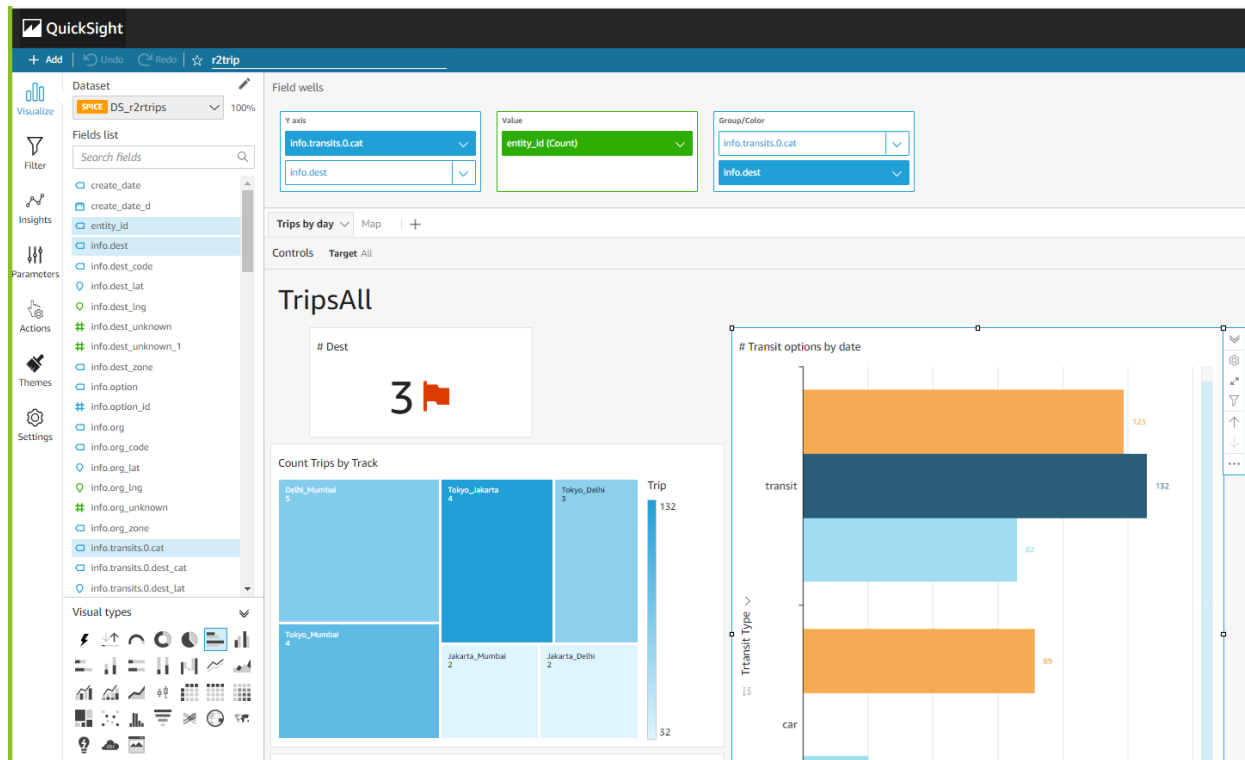
• Python - Web Scrapping

	cat	name	org_name	dest_name	duration	dist_transit	price_unit	min_price_unit
options								
1	0	transit	train	Tel Aviv Center	נתב"ג	1140.000000	11.000000	USD 3.000000
	1	walk	transfer	נתב"ג	Tel Aviv (TLV)	0.000000	0.000000	
	2	flight		Tel Aviv	Newark		9161.000000	USD 350.000000
	3	walk	transfer	Newark (EWR)	Terminal B	0.000000	0.000000	
	4	transit	tram	Terminal B	Newark Liberty Airport Station	360.000000	2.000000	USD 5.000000
	5	walk	transfer	Newark Liberty Airport Station	Newark Airport Railroad Station	0.000000	0.000000	
	6	transit	train	Newark Airport Railroad Station	New York Penn Station	1680.000000	18.000000	USD 13.000000
2	7	hotel		New York Penn Station				
	0	transit	train	Tel Aviv Center	נתב"ג	1140.000000	11.000000	USD 3
	1	walk	transfer	נתב"ג	Tel Aviv (TLV)	0.000000	0.000000	
	2	flight		Tel Aviv	New York JFK		9137.000000	USD 330
	3	walk	transfer	New York JFK (JFK)	Terminal 4	0.000000	0.000000	
	4	transit	tram	Terminal 4	Jamaica Station-Station D	720.000000	6.000000	
	5	walk	transfer	Jamaica Station-Station D	Jamaica	0.000000	0.000000	
3	6	transit	train	Jamaica	New York Penn Station	1200.000000	17.000000	USD 4
	7	hotel		New York Penn Station				
	0	transit	train	Tel Aviv Center	נתב"ג	1140.000000	11.000000	USD 3.000000
	1	walk	transfer	נתב"ג	Tel Aviv (TLV)	0.000000	0.000000	
	2	flight		Tel Aviv	New York La Guardia		9134.000000	USD 360.000000
	3	walk	transfer	New York La Guardia (LGA)	Terminal B/Hov Lane	0.000000	0.000000	
	4	transit	bus	Terminal B/Hov Lane	Woodside Av/61 St	1440.000000	4.000000	USD 1.000000
	5	walk	transfer	Woodside Av/61 St	Woodside	0.000000	0.000000	
	6	transit	train	Woodside	New York Penn Station	660.000000	8.000000	USD 3.000000
	7	hotel		New York Penn Station				
	0	transit	train	Tel Aviv Center	נתב"ג	1140	11.000000	USD 3.000000
	1	walk	transfer	נתב"ג	Tel Aviv (TLV)	0	0.000000	
	2	flight		Tel Aviv	Philadelphia		9288.000000	USD 280.000000
	3	walk	walk	Philadelphia (PHL)	Airport Terminal A	499.256	0.000000	



QuickSight

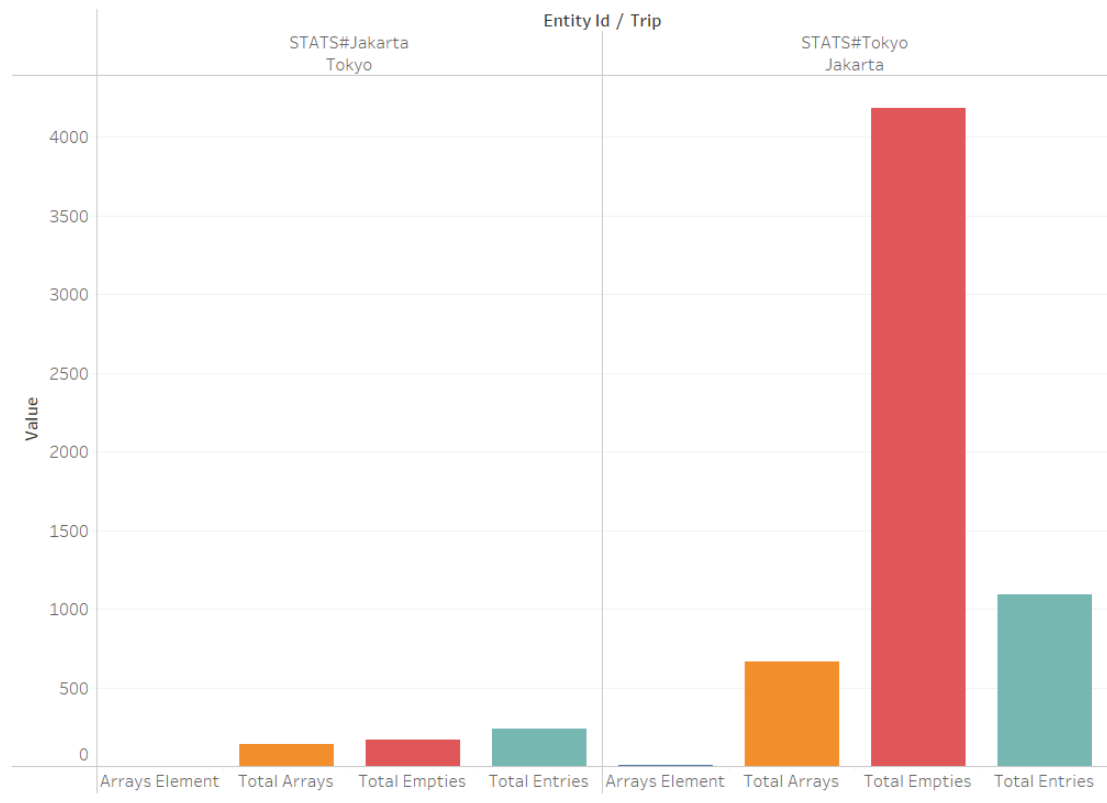
- Create analytics and display Dashboards



Tableau

- Create Reports

trips - hierarchy elemnts distributions



Thank you