ENGINEERING PROJECT MVP:

PROJECT TASK (HIGH LEVEL): TO BUILD A DATA PIPELINE

SPECIFIC PROJECT CHOICE: TO BUILD A DASH BOARD USING FLIGHT DATA

My project choice is to feed flight data into a pipeline that respond to customer queries. Customers can ask a wide range of questions but we are limiting ourselves to flight_related questions.

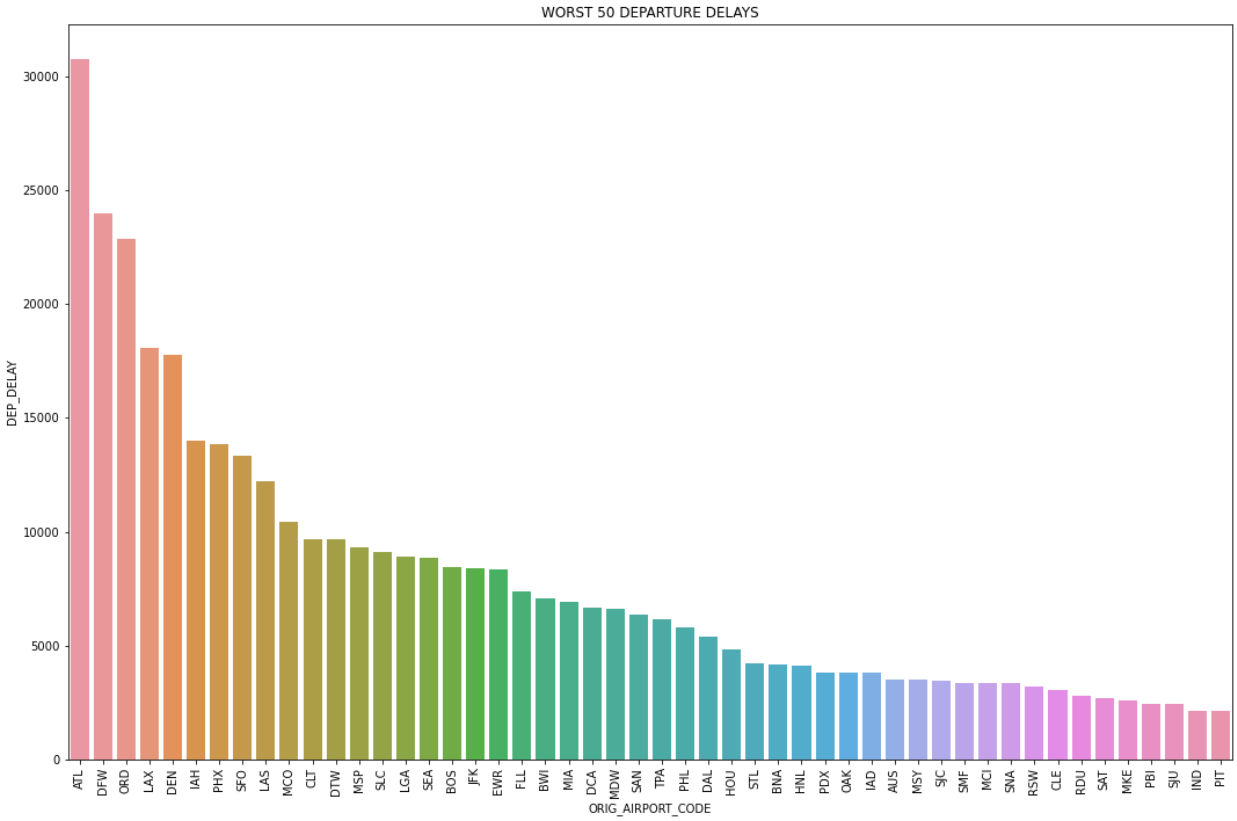
MVP:

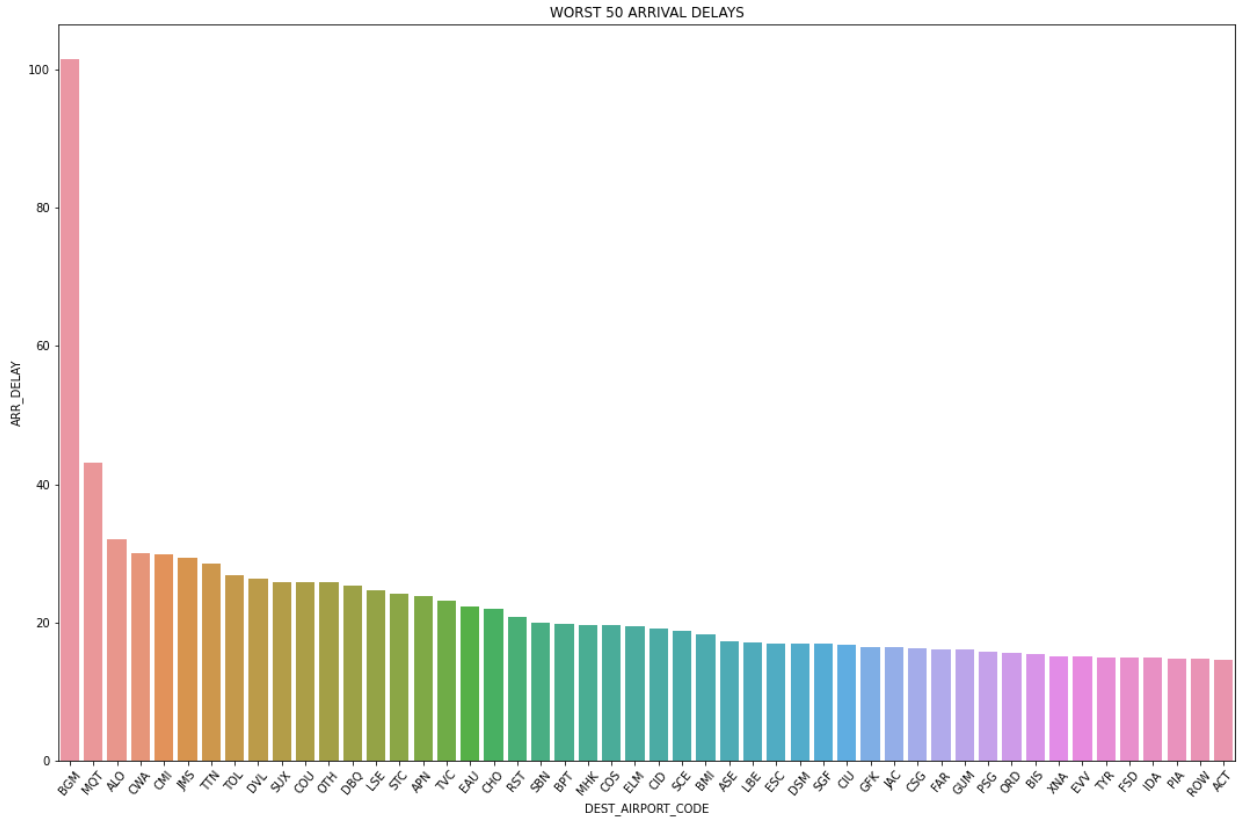
MVP target was a working model, responding to basic queries, for the customer to pass simple delay-related queries. Unfortunately, we have had to:

Having a clear dataset, with sample plots /sample analysis of what client may expect. 500,000 records min.).

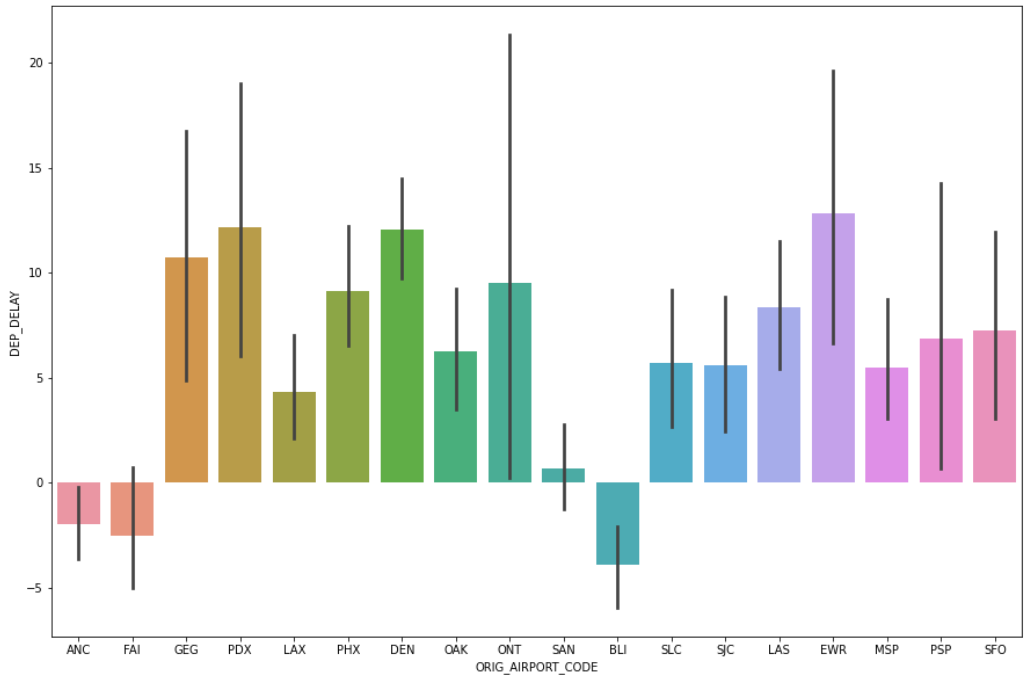
Front-end will be powered by Streamlit.

Major Challenges: MacM1-Tablearu compatibility issues issues. Hence the shift from Tableau to Streamlit. Below are basic plots of the top 50 airports with the largest delays as well as the top 50 arrival-delay-prone airports. At least you know what to avoid.


WORST 50 DEPARTURE DELAYS
DEP_DELAY
0
5000
10000
15000
20000
25000
30000
ATL
DFW
ORD
LAX
DEN
IAH
PHX
SFO
LAS
MCO
CLT
DTW
MSP
SLC
LGA
SEA
BOS
JFK
EWR
FLL
BWI
MIA
DCA
MDW
SAN
TPA
PHL
DAL
HOU
STL
BNA
HNL
PDX
OAK
IAD
AUS
MSY
SJC
SMF
MCI
SNA
RSW
CLE
RDU
SAT
MKE
PBI
SJU
IND
PIT
ORIG_AIRPORT_CODE


WORST 50 ARRIVAL DELAYS
ARR_DELAY
0
20
40
60
80
100
BGM
MQT
ALO
CWA
CMI
JMS
TTN
TOL
DVL
SUX
COU
OTH
DBQ
LSE
STC
APN
TVC
EAU
CHO
RST
SBN
BPT
MHK
COS
ELM
CID
SCE
BMI
ASE
LBE
ESC
DSM
SGF
CIU
GFK
IAC
CSG
FAR
GUM
PSG
ORD
BIS
XNA
EVV
TYR
FSD
IDA
PIA
ROW
ACT
DEST_AIRPORT_CODE

RANDOM DEPARTURE DELAY DISTRIBUTION



Dennis Ssekamaanya