ARC_Direct

Data Assessment

Team: Shu-Ping Chen, Jinping Guo,
Jing Lin, Shih-Siang Lin, Po-Han Yen

Business Problem

An increase in direct booking from airlines will bring about a huge loss to both ARC itself and the cooperative travel agencies.

Objective

Forecast the direct booking from airlines during 2022 - 2024 to help ARC understand the big picture of post-covid air travel and come up with successful business strategies in advance.

Data Source

Internal data source:

- ARC Snowflake:
  Issue year, issue week, carrier name, origin country, destination country, advanced purchased days, transaction type, trip type, cabin class, tickets count

141 million rows and 10 columns (10GB)

External data source:

- COVID-19: Stringency Index
  - https://ourworldindata.org/covid-stringency-index
- Jet Fuel Spot Price
  - https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EER_EPJK_PF4_RGC_DPG&f=W

Unavailable Data

- Business or leisure traveler
- Fuel price & covid policy for the remaining of 2022 to the end of 2024
- Domestic vs International covid travel policies for certain countries

Data Cleansing, Preparation

Utilize SQL to extract target features:

IISSUE_YR, ISSUE_WEEK, TRAVELER_TYPE, GEO_SEGMENTS, TOTAL_TICKETS
2017 to 2022 week 18, Data Source = Direct, Transaction type = Issue & Exchange
(The most relevant and completed data)

Transfer tickets from week 53 to week 52 in 2020 & 2021:

2017-2019 have only 52 weeks and 2020-2021 have 53 weeks with the last week containing only a few days

Extract fuel price by year and week

Extract covid policy index for top 10 countries
Calculate average covid policy index for all countries and for remaining countries

Combine fuel price and covid policy with transactions of corresponding countries by year and week

Feature Engineering

Time

Combine year and week to generate the feature “time” which is more suitable for time series models (Constant increment format)

Traveler_type

Classify travelers to business and leisure based on different cabin classes
(Business: First & Business class, Leisure: Other classes)

Geo_segments

Top 10 countries with the most ticket amounts
Other countries with fewer ticket amounts
Domestic vs International

Top 10 countries have significant impacts on total tickets

	DESTINATION_COUNTRY	TOTAL_TICKETS
1	US	1,648,664,071
2	CN	766,016,190
3	BR	198,937,142
4	CA	144,070,615
5	DE	143,668,167
6	RU	142,067,113
7	AU	130,249,392
8	GB	129,924,998
9	FR	110,276,013
10	TR	107,338,559
11	KR	91,418,445
12	ES	85,664,157
13	IT	74,493,531
14	SA	72,051,631
15	CO	61,043,323
16	IN	55,035,977
17	JP	43,475,533
18	MX	43,231,840
19	AE	39,907,846

Model Considerations

Forecast of fuel price using unsupervised ML methods (Abandoned: Time and resource restrictions)

Reference public resources and reliable external predictions

Linear Regression with potential nonlinear transformations

---Forecast of covid policy with current trend by country.

SARIMAX & Ensemble methods

---Forecast total tickets with seasonality, trend and exogenous variables etc.

---Ensemble predictions from individual segments for better forecasting.

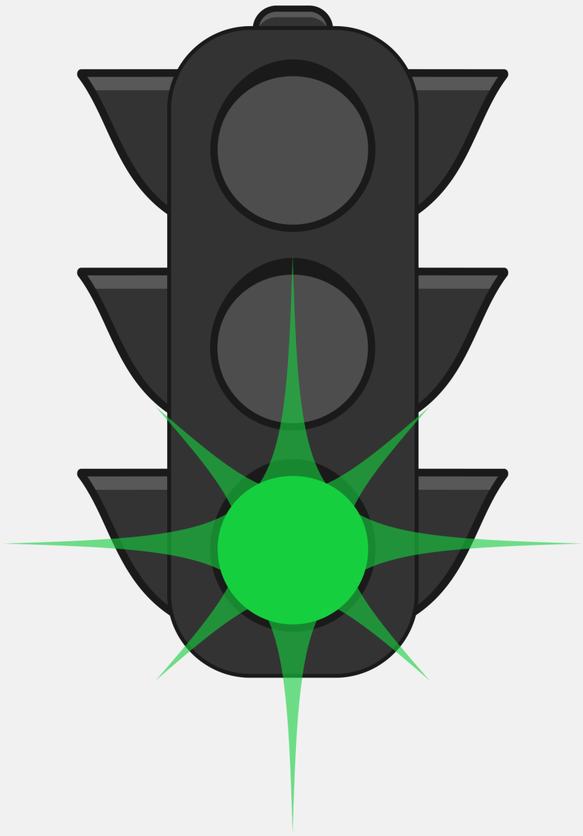
Preliminary Findings

Pre-covid: Obvious seasonality & trend

Post-covid: Non-obvious seasonality & trend

Indicating for a potential two-part analysis

Project Success Assessment



We have successfully narrowed down the scope of the business problem to a handleable situation by extracting and aggregating necessary features. However, we still keep our analysis to the highest standard by introducing sophisticated segmentations and incorporating covid policies and fuel prices, which we believe are most relevant to post-covid ticket forecast.

We have cleaned and transformed our data to appropriate formats for training time series machine learning models. We are well prepared to implement our segmentation analysis and validate the results to refine our forecast.

Thank You

ARC

UNIVERSITY OF MARYLAND 18 56

ROBERT H. SMITH SCHOOL OF BUSINESS