Time Series Summer 2020

Forecasting International Flight Passengers

Lola Johnston | Julian Kleindiek | Kelley Monzella | Jerry Sha

AGENDA

Our goal is to forecast the number of travelers flying to different regions of the world from America over time.

PROBLEM TO SOLVE

Can we predict how many people will fly to specific regions of the world each month, over the next year?

BUSINESS APPLICATION

- Airlines logistics
- Tourism planning

Our dataset comprises international flights over three decades.

DATA SOURCE

U.S. Department of Transportation Office of the Assistant Secretary for Aviation and International Affairs

MONTHLY PASSENGERS

600K+ ROWS

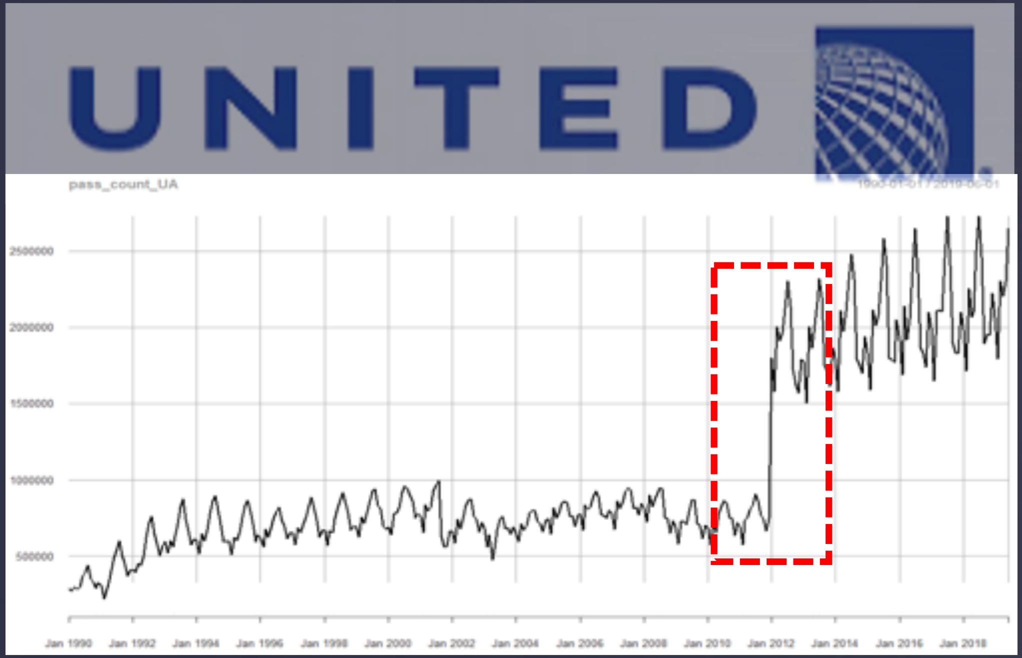
YEARS 1990 - 2019

550+ airlines
900+ airports (non-US)
160+ countries

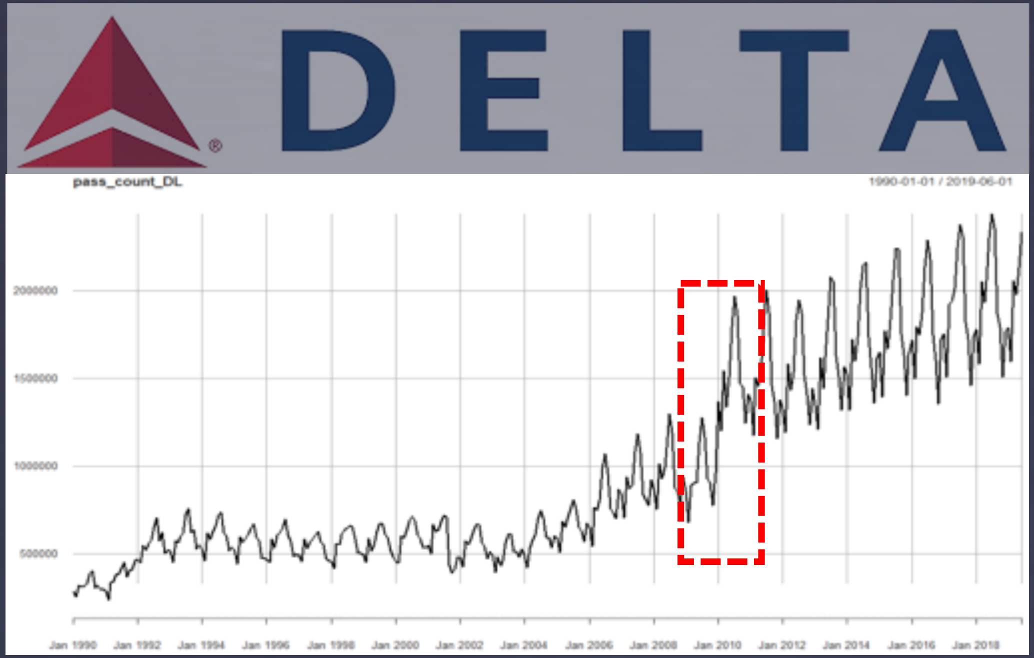
PART 2/6

EDA

We attempted aggregation by carrier but found the data cumbersome due to frequent mergers between airlines.



United Airlines
(merged with Continental Airlines in 2012)



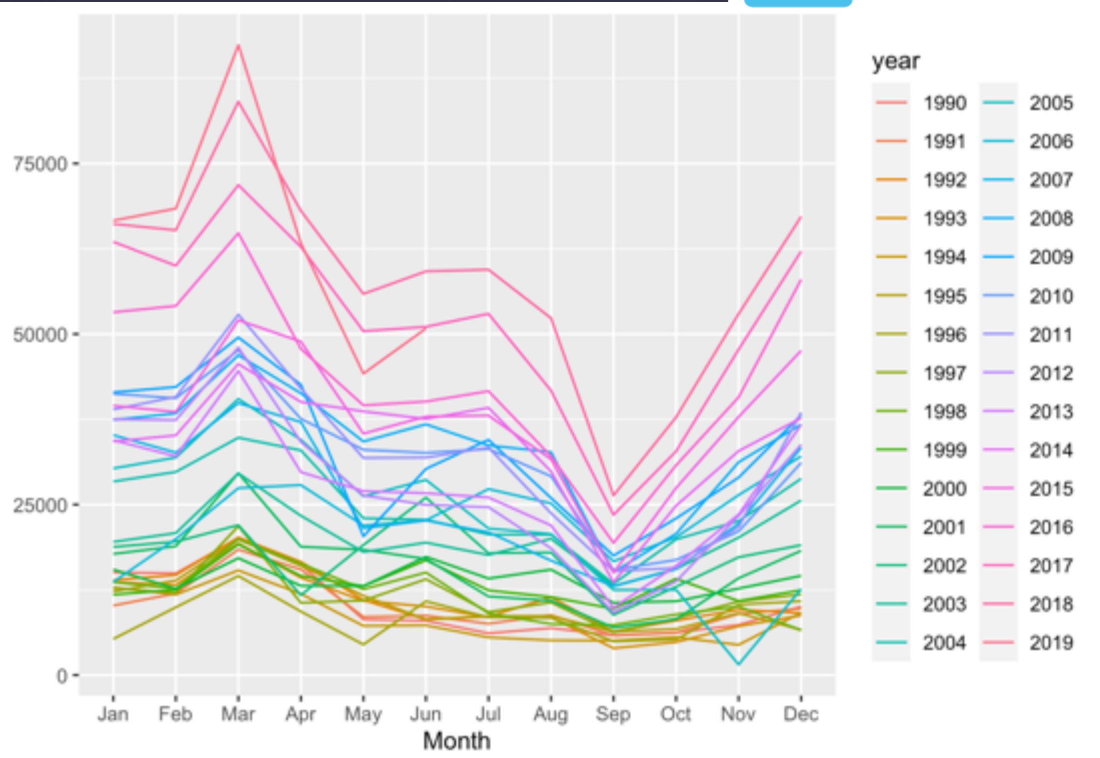
Delta Airlines
(merged with Northwest Airlines in 2010)

We explored our data by airports but didn't consider it further because of the number of possible combinations.

CHICAGO

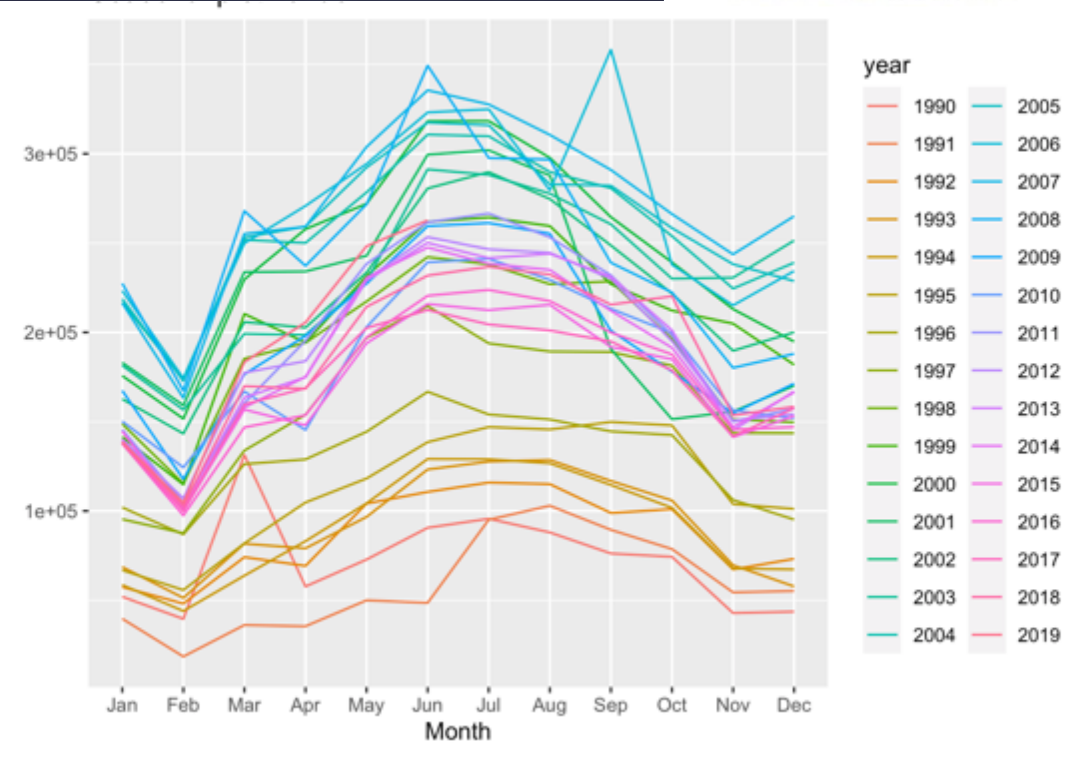
Seasonal plot ORD-CUN

CANCUN INTERNATIONAL AIRPORT



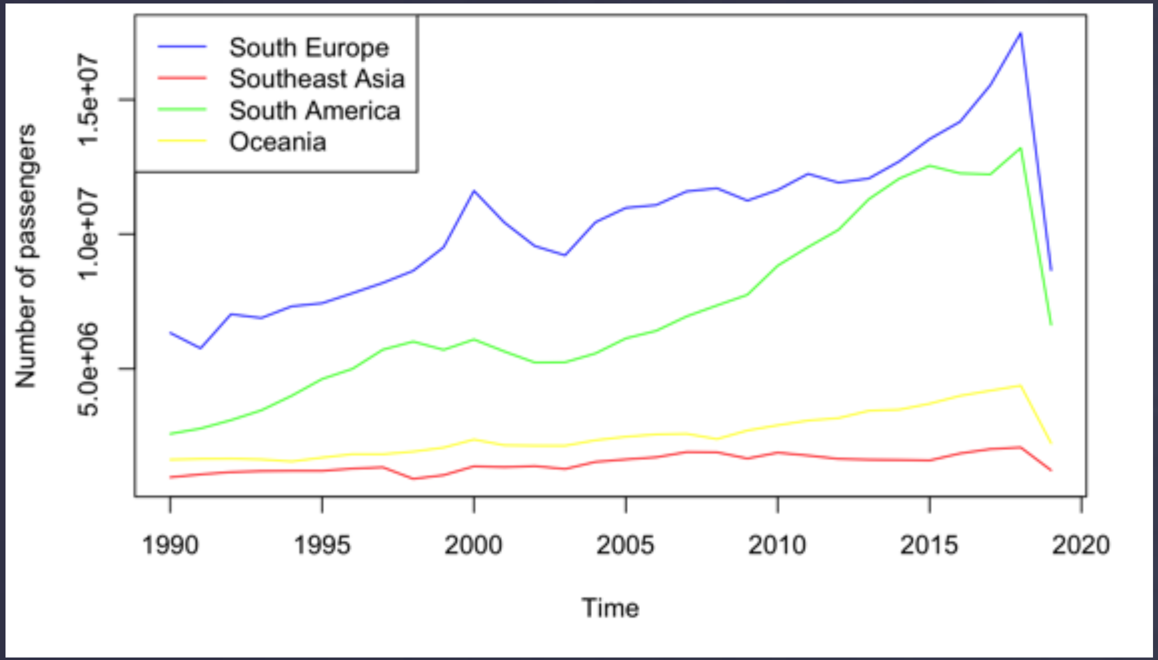
Seasonal plot ORD-LHR

Heathrow

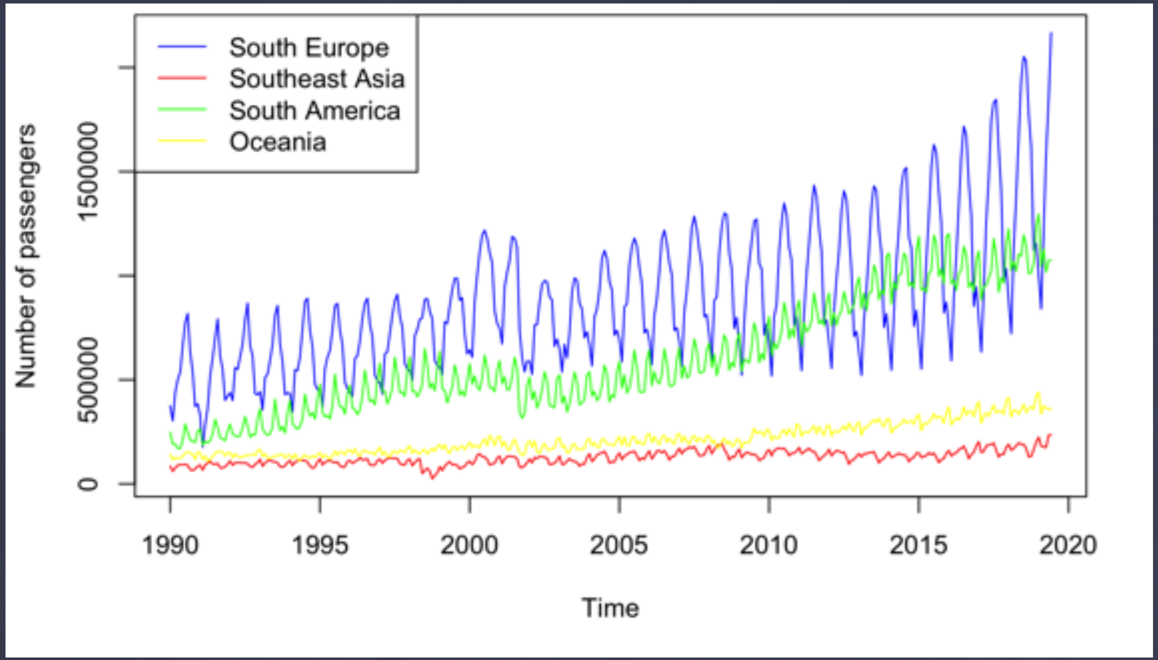


We then explored our data by region and selected South Europe as the focus for our modeling efforts.

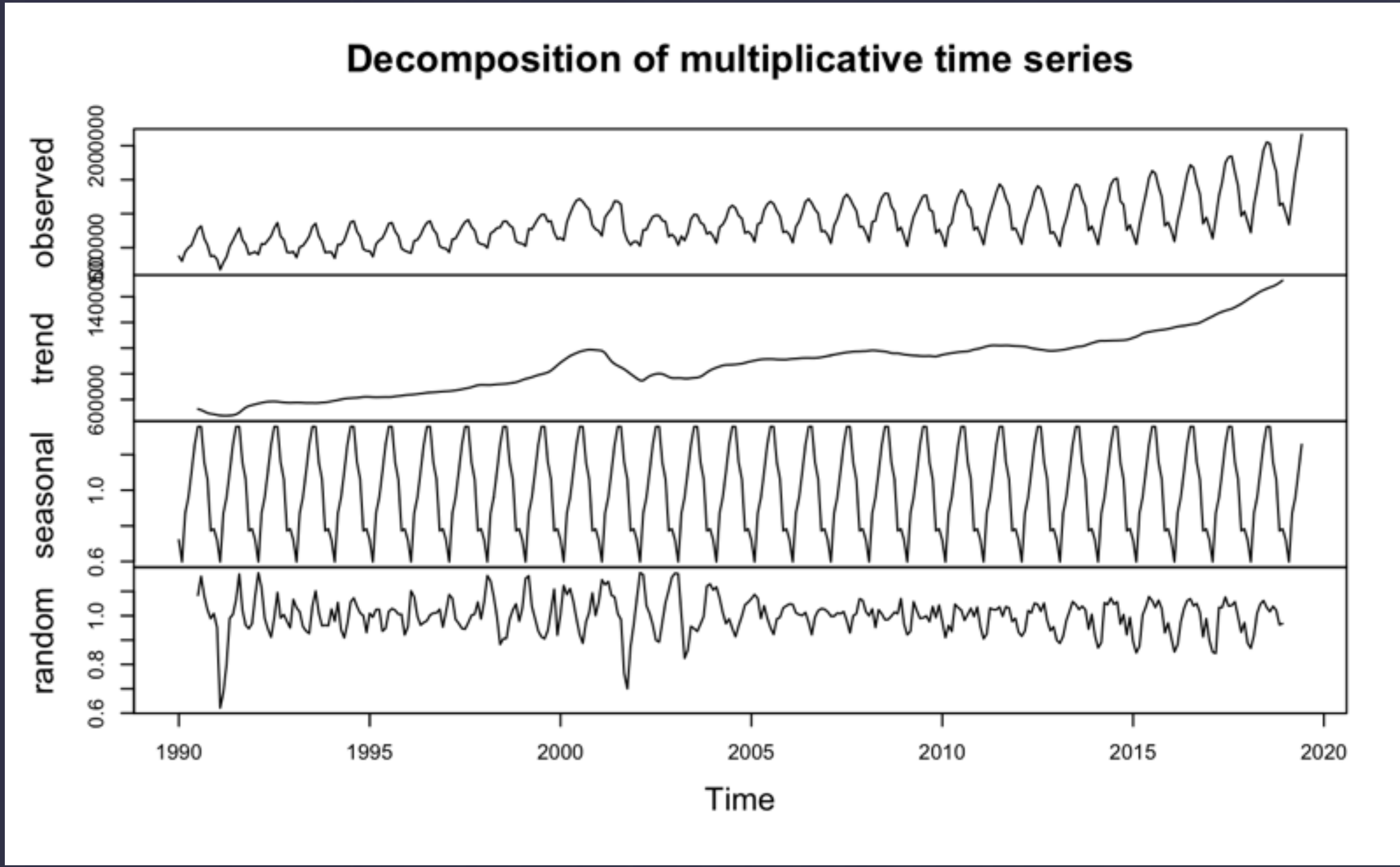
Int'l travel from the US by **year**



Int'l travel from the US by **month**



South Europe represented an interesting subset of our data for further analysis.



1) **Positive trend** validated by KPSS test with a shock in 2001

2) **Multiplicative** time series so Box-Cox transformation needed

3) **Multiple seasonality**

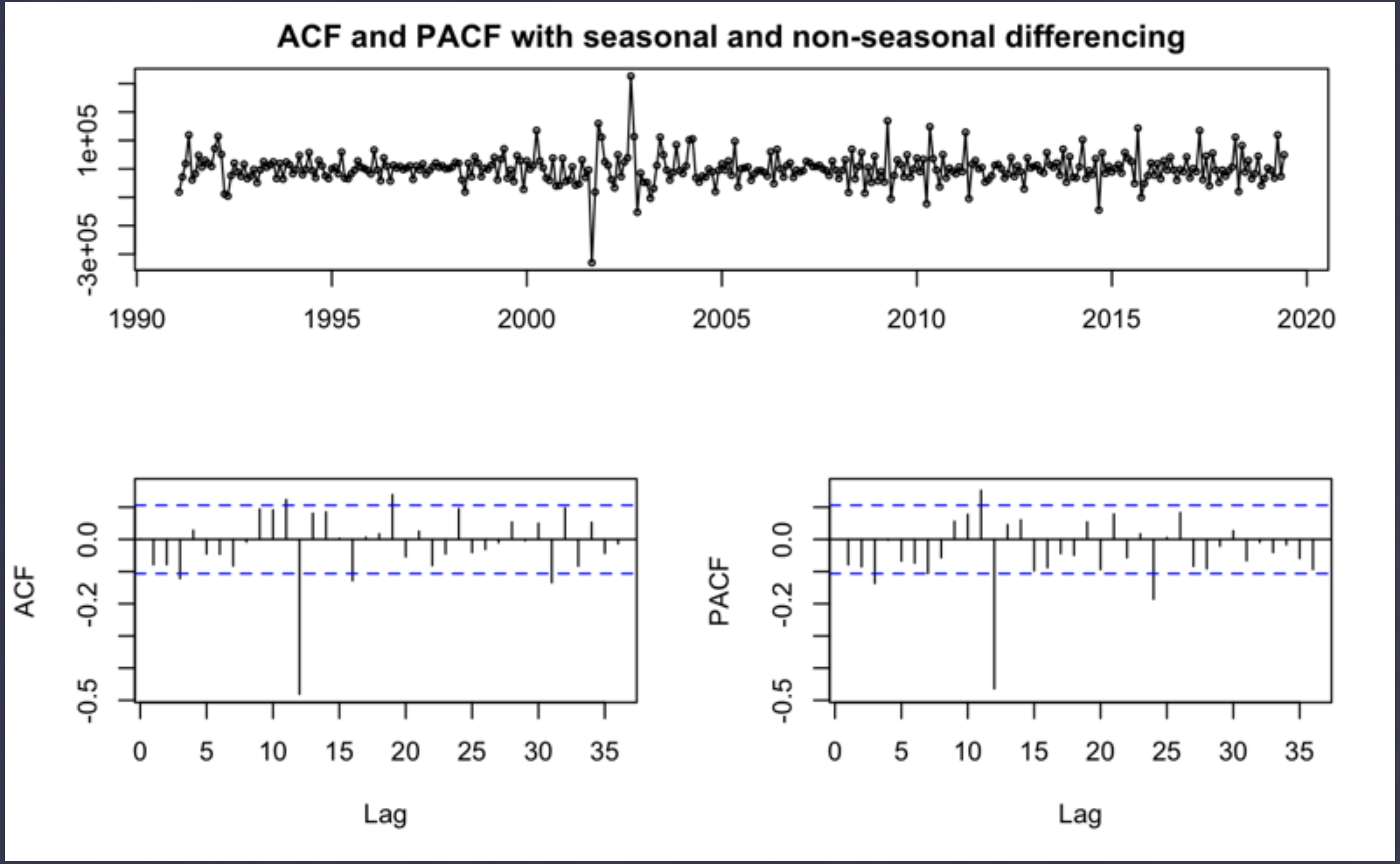
The ACF and PACF helped us better understand the data for modeling.

Seasonal differencing
From earlier plots, we know that there was seasonality in the data (D = 1; s = 12)

Non-seasonal differencing
From earlier plots, we know that the data was non-stationary (d = 1)

AFC
Cutoff at lag 12
(Q could be 1)

PACF
Cutoff at lag 12 and 24
(P could be 1 or 2)



PART 3/6

Model selection

MODELS

AVERAGE

NAÏVE

SEASONAL NAÏVE

RANDOM WALK

SES

HOLT WINTERS

ETS

TBATS

ARIMA

AUTO ARIMA

VAR

MODELS	QUICK DESCRIPTION
AVERAGE	Average of historical data, flat forecast
NAÏVE	Last point, flat forecast
SEASONAL NAÏVE	Last seasonal period
RANDOM WALK	Last point but with a constant/drift
SES	Considers level only, not suitable for trend & seasonality
HOLT WINTERS	SES but allows trend and seasonality (only 1 seasonal)
ETS	Holt Winters w/ state space, accepts non-stationary data
TBATS	Accepts non- stationary data, Holy grail
ARIMA	Requires stationary data
AUTO ARIMA	Accepts non-stationary data, Holy grail
VAR	Multivariate

We learned that inconsistencies across inputs made evaluations difficult.

problems

Insignificant insights to guide model selection

Inconsistency across inputs

We learned that inconsistencies across inputs made evaluations difficult.

problems

Insignificant insights to guide model selection

Inconsistency across inputs

solution

Use consistent input across all models:

differenced data

Final models.

MODELS	QUICK DESCRIPTION	PARAMETERS
ETS	Holt Winters w/ state space, accepts non-stationary data	(M,Ad,M) α 0.16 | β 0.09 | γ 0.84 | ϕ 0.85
TBATS	Accepts non- stationary data, Holy grail	(0.001, {0,0}, -, {<12,5>}) λ 0.0005 | α 0.77 | γ 1: 0.001 | γ 2: 0.011 fourier 5
ARIMA	Requires stationary data	(3,1,3)(1,1,2)[12] boxcox lambda= -0.002
AUTO ARIMA	Accepts non-stationary data, Holy grail	(1,1,0)(2,1,0)[12] boxcox lambda= -0.002
VAR	Multivariate	VAR 4 , Additional variables: CPI & Inflation

Using VAR() we attempted multi-variate forecasting.

PRICE ELASTICITY OF FLIGHTS = -0.7

⬆ PRICE +10%

⬇ DEMAND - 7%

Hospitality Research Journal, Volume: 17 issue: 3, page(s): 27-39
Issue published: May 1, 1994
Geoffrey I. CrouchWorld Tourism Education and Research Centre The University of Calgary

Using VAR() we attempted multi-variate forecasting.

PRICE ELASTICITY OF FLIGHTS = -0.7

PRICE +10%

DEMAND - 7%

ADDITIONAL VARIABLES:

CPI

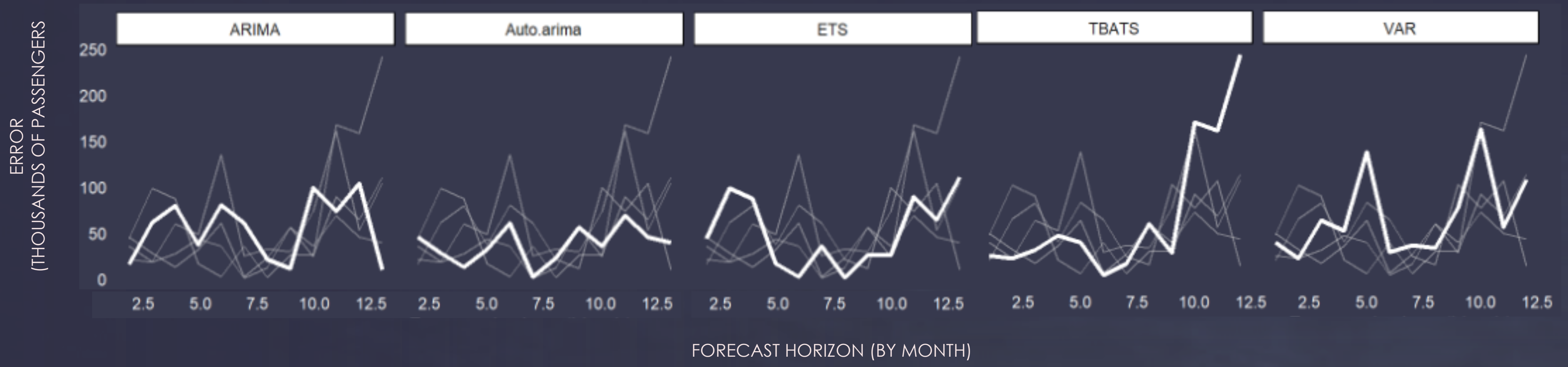
INFLATION

Hospitality Research Journal, Volume: 17 issue: 3, page(s): 27-39
Issue published: May 1, 1994
Geoffrey I. CrouchWorld Tourism Education and Research Centre The University of Calgary

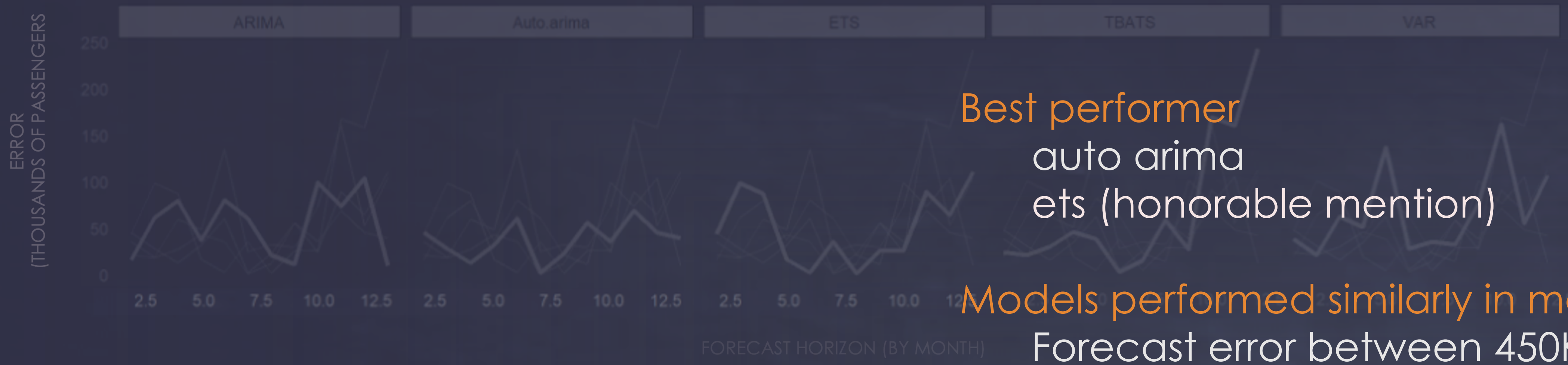
PART 4/6

Evaluation

Forecast accuracy by month



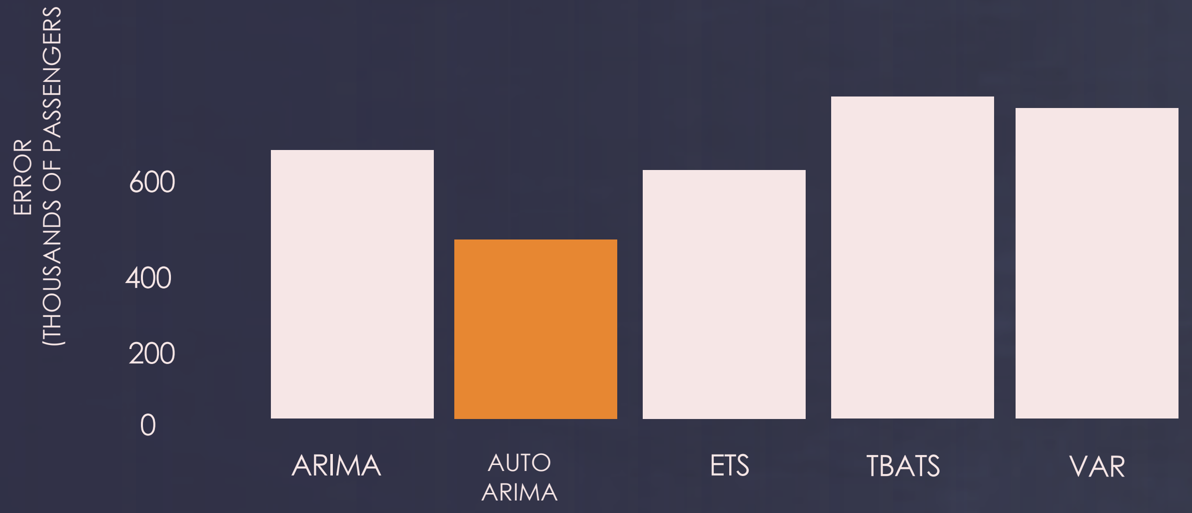
Forecast accuracy takeaways



Best performer
- auto arima
- ets (honorable mention)

Models performed similarly in many ways
- Forecast error between 450K and 800K
- Multiple seasonality

Notable difference exist
- TBATS forecast horizon
- Auto.arima vs. hand-tuned
- Not shown: naïve, average, other poor performing models



PART 5/6

Implementation

Future directions.

International flights forecasting - Traveling to Southern Europe

Choose your parameters and start modeling time series data.

Departure airport

EWR

Arrival airport

BCN

Time Series Model

Autoarima

Choose forecast horizon

1 6 12

1 3 5 7 9 11 12

Submit

Step 1. Plot your data | Step 2. Transform your data | Step 3. Forecast your data | Step 4. Check your residuals

Stationarity Summary:

Your data are not stationary according to a KPSS test and your data are not stationary according to an ADF test.

KPSS summary table

```
        KPSS Test for Level Stationarity

data:  ts(selectedTSData())
KPSS Level = 2.335, Truncation lag parameter = 2, p-value = 0.01
```

ADF summary table

```
        Augmented Dickey-Fuller Test

data:  ts(selectedTSData())
Dickey-Fuller = -1.4801, Lag order = 5, p-value = 0.7933
alternative hypothesis: stationary
```

Box Cox Transformation Summary:


takeaways
PART 6/6

We successfully achieved our objective of forecasting airline passenger count by region.

CONCLUSION

- International flight data is well-suited for time series modeling
- More complicated models don't always deliver superior results

FUTURE WORK

- Introduce additional predictors (weather, economics, etc.)
- Additional model frameworks (NNAR, RNN, ARCH/GARCH)

Thank you.