Let’s Take a Trip…

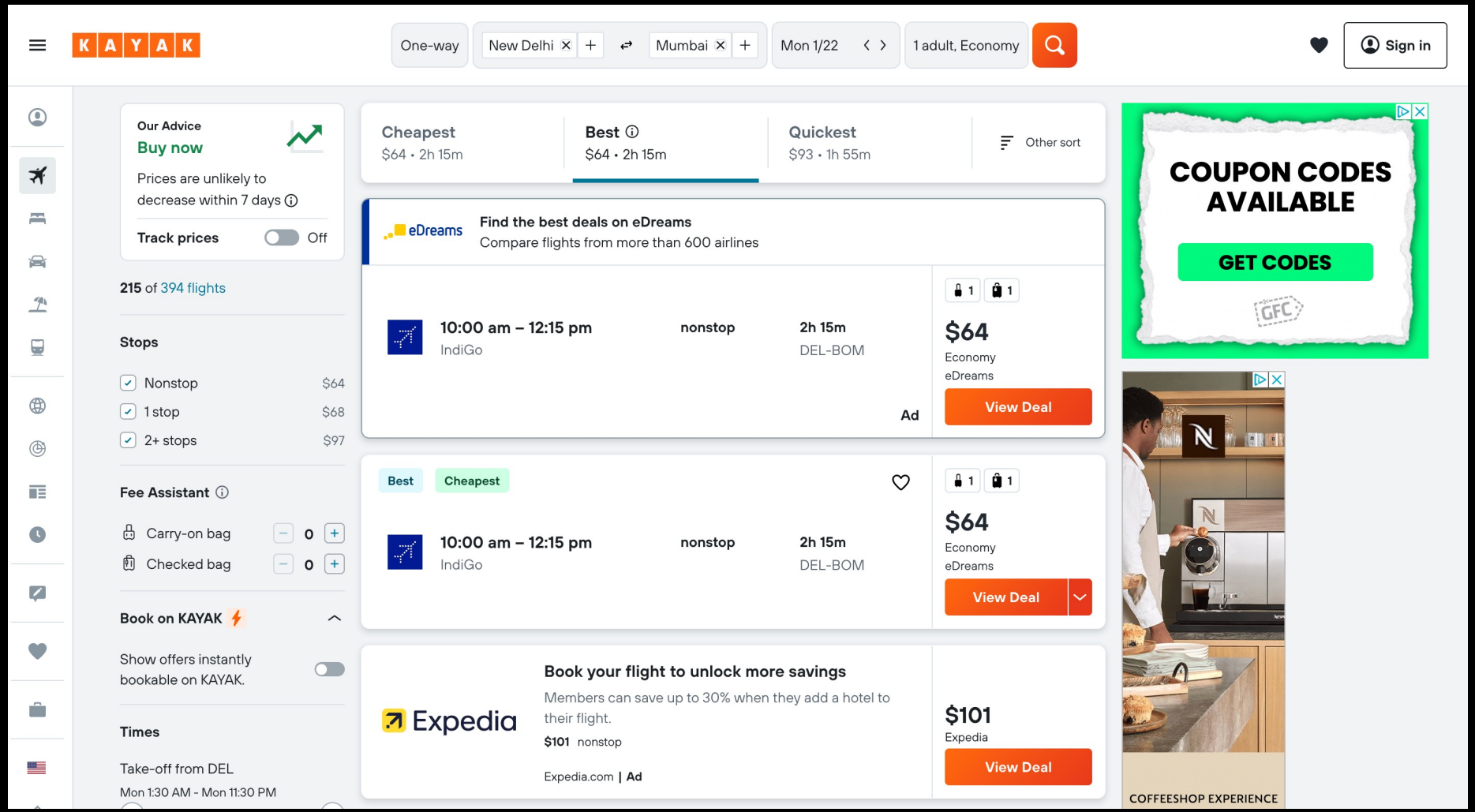
By Morgan Snellgrove

To India!

We’re planning to go to

- Delhi
- Mumbai
- Bangalore
- Kolkata
- Hyderabad

Are these good tickets?

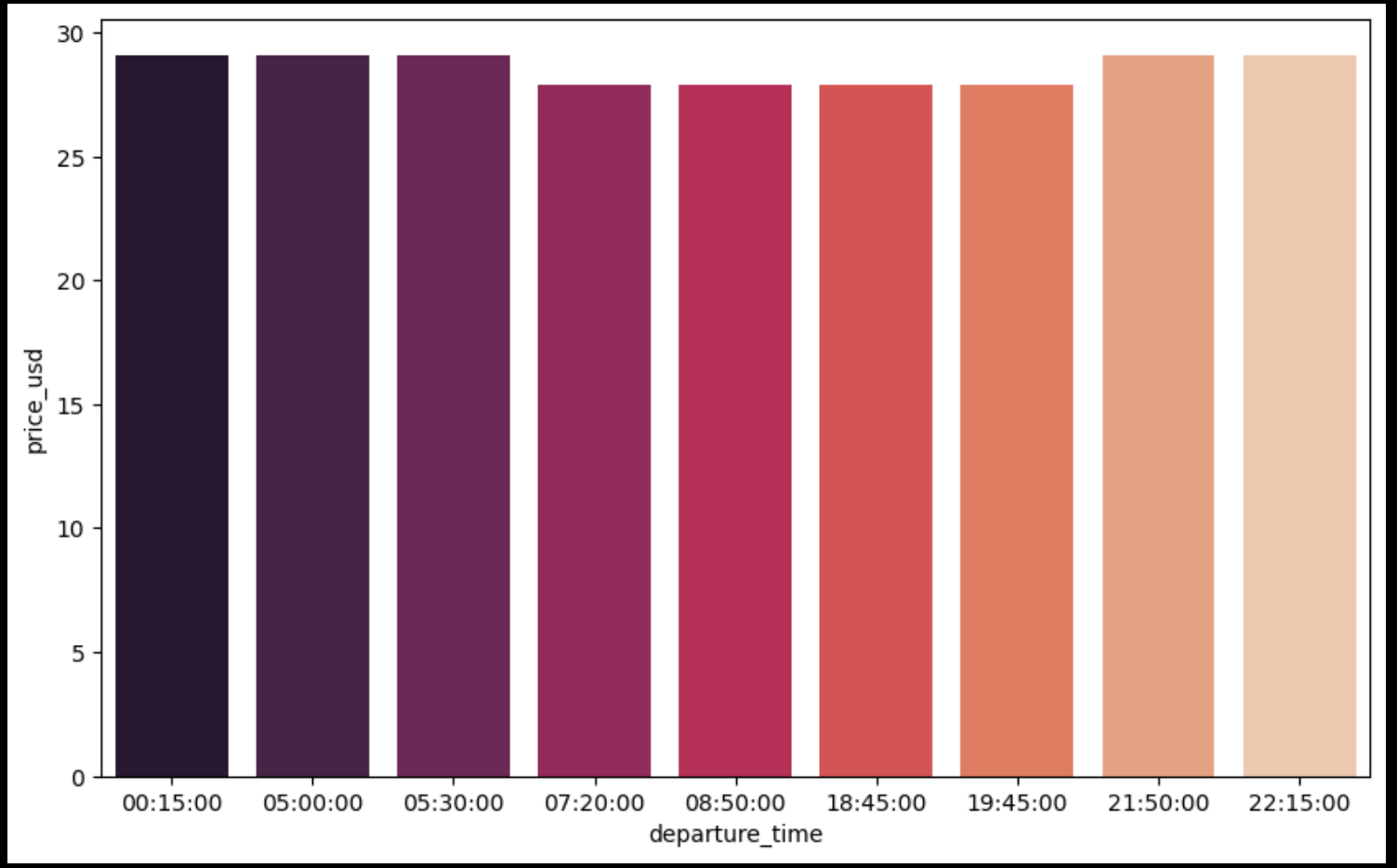


Data to the rescue!

- 300,000 examples!
- What is the cheapest this ticket has been?
- What is the most affordable airline company?
- How far in advance should I purchase?
- What day of the week has the best ticket prices?

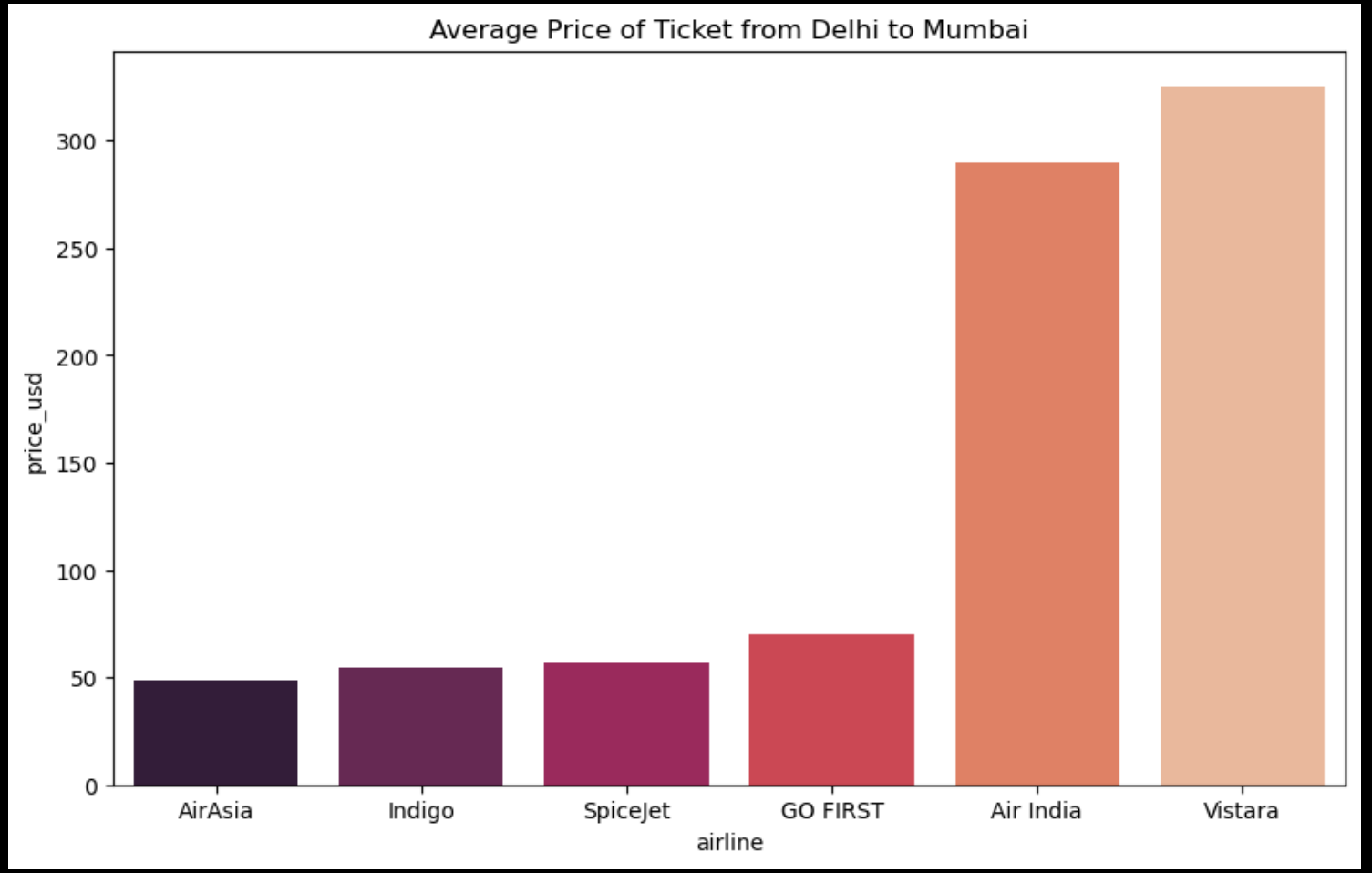
Delhi to Mumbai

- Less than $30!
- Middle of the night is more expensive
- 8:50am departure time



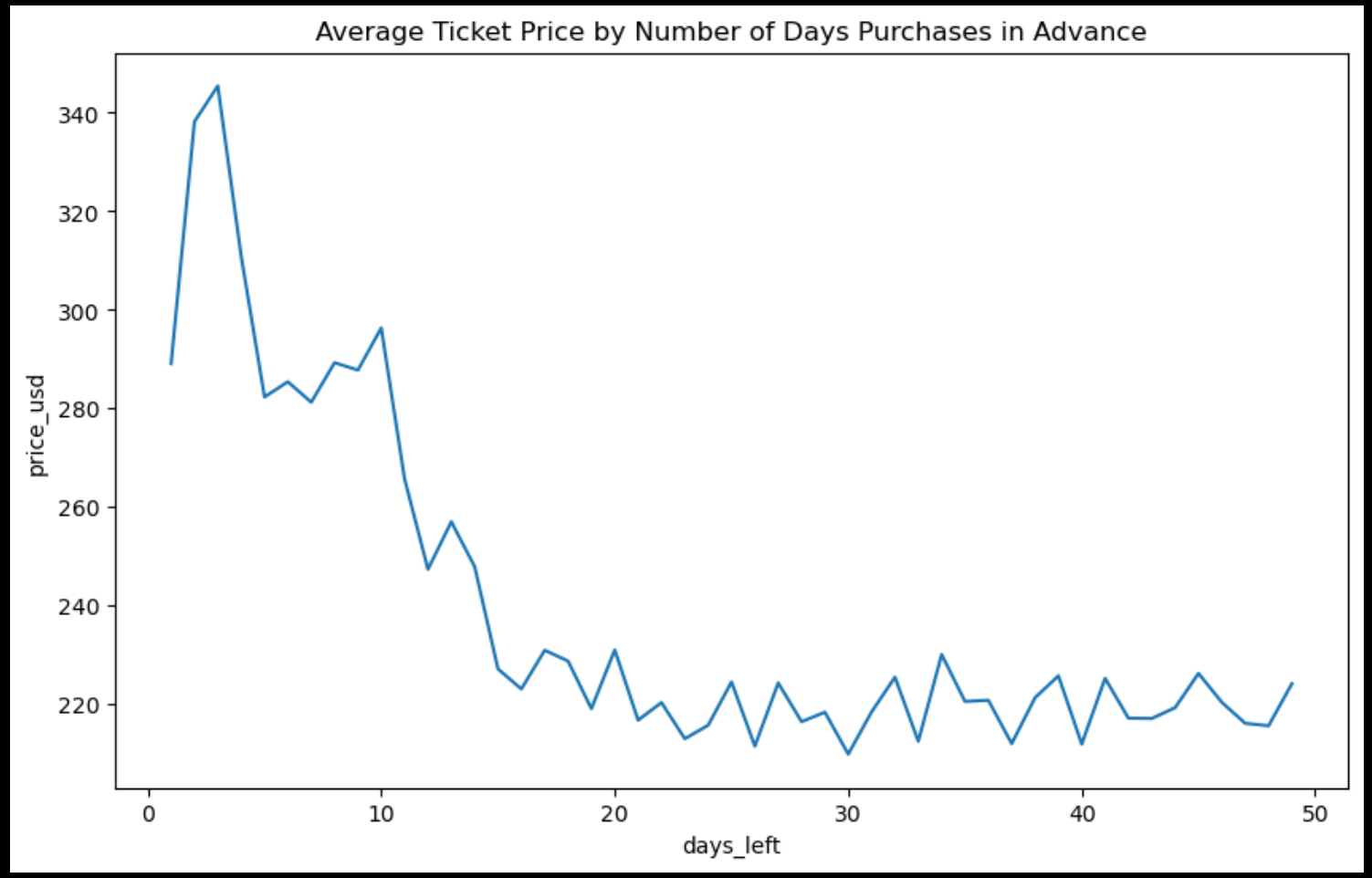
Most Affordable Airline Company

- AirAsia
- Cheapest tickets were from SpiceJet
- Main difference is departure date: March 29th



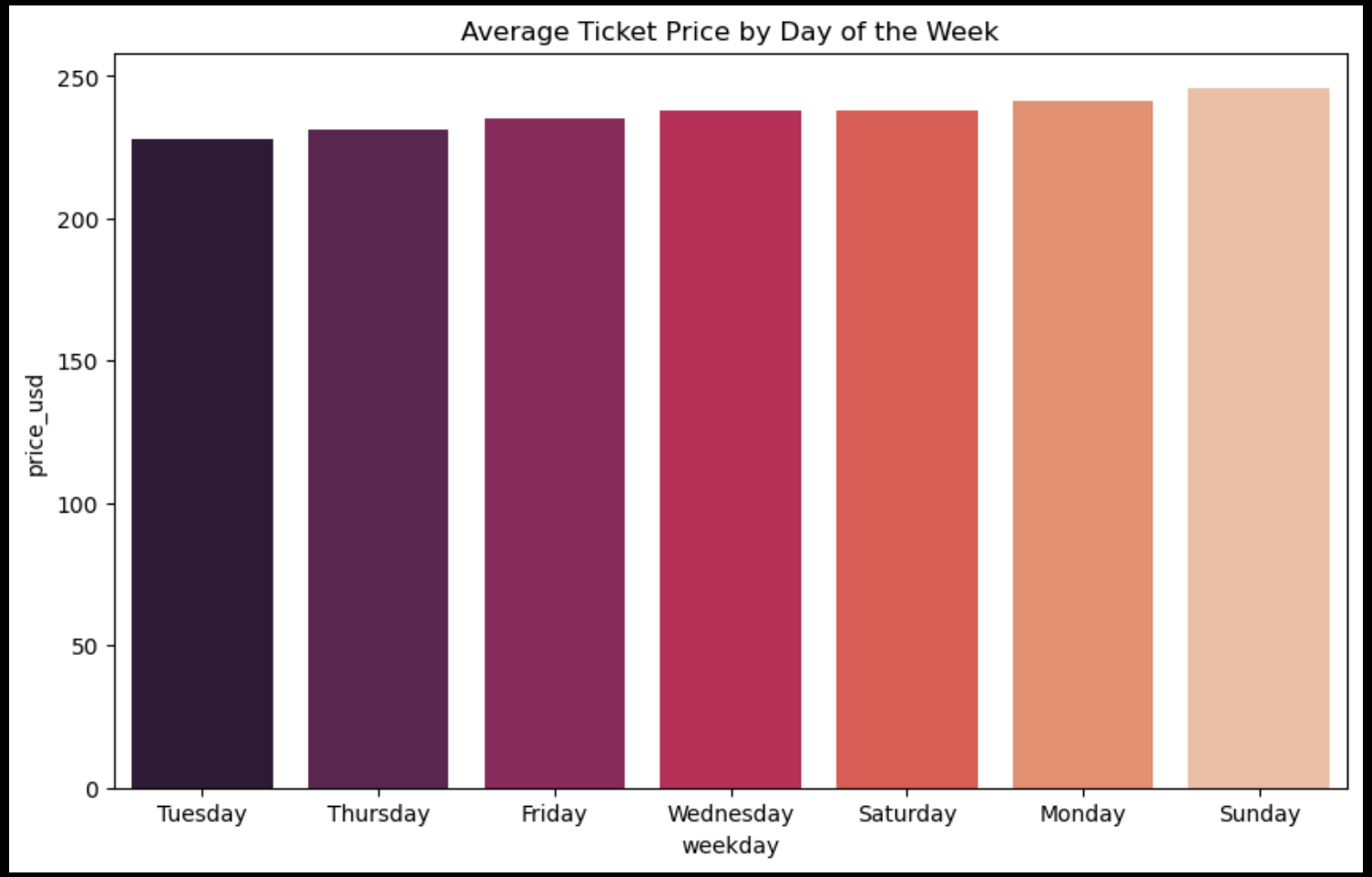
Does it matter when I buy my ticket?

Tickets purchased 20 days or less before departure cost an average of 22% more!



What day of the week is best to fly?

Tuesday is the cheapest on average!



So far we know…

1. We can find tickets as cheap as $30.
2. We need to check prices at AirAsia.
3. We need to purchase more than 20 days in advance.
4. Flying on Tuesdays might save us some money.

What else can we do with this dataset?

Can we predict ticket prices?

Absolutely!

Preparing our data:

Issues:

- Dates
- Times
- Categorical Features
- Huge scale differences

Fixes:

- Extract month and day
- Extract hour and minute
- Encode using binary indicators
- Apply Standard Scaler

Selecting a Model

Now that our data is prepped, here's what we're working with:

- We've split our data: 75% training, 25% testing
- Our target is 'price_usd'
- Dummy Regressor for baseline comparison: strategy = mean

Test Set Scores	Dummy Regressor	Linear	Ridge	Lasso	Random Forest	Tree-based XGBoost	Linear-based XGBoost
R - Squared	-4.69	0.91	0.925	0.92	0.99	0.98	0.92
MAE	$241.66	$55.64	$51.68	$51.62	$10.02	$24.25	$51.41
RMSE	$277.54	$84.90	$76.10	$78.47	$26.69	$41.68	$76.24

Random Forest Model

- Features account for 99% of the variance in price usd.
- MAE reduced by over 95%
- RMSE reduced by over 90%

Test Set Scores	Dummy Regressor	Random Forest Model
R-Squared	-4.69	0.99
MAE	$241.66	$10.02
RMSE	$277.54	$26.69

With this data, we can…

Know the cheapest ticket price

Know which airline is most affordable

Learn how far in advance to purchase your tickets

Know which day of the week is cheapest to fly

Predict ticket prices with surprising accuracy