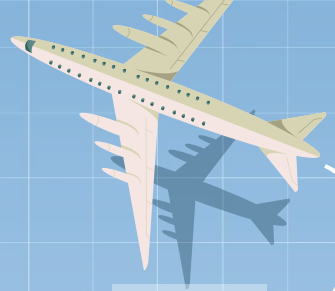
SPARK WINGS

By Benoy Thomas, Navraj Singh, Luis Jaco

TABLE OF CONTENTS



HOW ARE BAGS LOADED

Bulk Loaded

- Bags are loaded one by one
- Very flexible and accommodate varying cargo sizes
- Slow loading/offloading
- Less efficient space utilization

Unit Load Devices (ULDs)

- Bags are loaded into specific ULDs
- Better space optimization
- # of ULDs are decided by Turnaround Coordinator (TCO) T-90 minutes before departure
- ULDs add extra weight and directly impact quantity of freight cargo

HOW ARE MEALS LOADED

- Meals are ordered T-4 Hours before departure
- Done manually and can't be changed after T-2 Hours before departure
- Usually based off the current booking, and a buffer is put in place which causes inefficiency and food wastage.
- Only exception are special meals (Gluten/Halal/Kosher) but rest of meals are ordered based off cabin (Upper J, Premium W, Economy Y).

G-VSRB Airbus A330-900neo

Economy	Premium	Upper	Cargo
185 Seats 1Bag/23kg	46 Seats 2Bag/23kg	32 Seats 2Bag/32kg	16,000 Kilos

PROBLEM STATEMENT

Cargo Space Efficiency

- Airlines prioritize revenue passengers and baggage over cargo
- Turnaround Coordinators often overestimate Passenger Numbers / Bag Count
- Which results in, reduced cargo space availability, Increased aircraft carbon footprint, Overall efficiency loss

Challenges in Flight Meal Estimation

- Airlines struggle to accurately estimate number of meals per flight
- Overestimating leads to:
  - Increased food waste
  - Higher operational costs

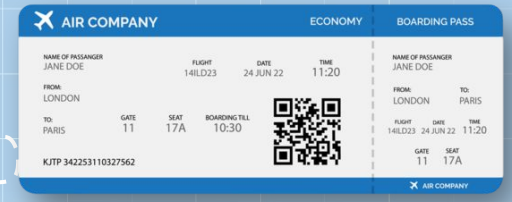
Project Goal

- Develop a machine learning model to predict final passenger count to optimize meal quantity and maximize cargo transport.
- Leverage historical data from Virgin Atlantic flights, including:
  - Date
  - Passenger count
  - Booking class
  - Time of day & day of the week
  - Seasonal trends
  - Special assistance

OUR MOTIVATION

NAV

Nav **works at Virgin JFK**, and deals with this problem everyday. See's food being wasted and a lot of inefficiency everyday.



LUIS

Luis is a **avid traveler**, but enthusiastic about **sustainability** and always tries to book flights with the lowest emissions.

BENOY

Benoy l**oves traveling** and passionate about **applying data science skills** so, this is a perfect mix of his two interests.

DATA SOURCE & CLEANING

Manual Data Collection by Operations Agents

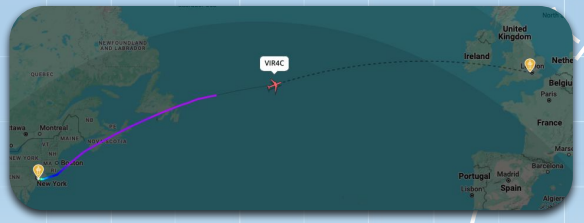
- Daily logs from JFK Terminal 4 for Virgin Atlantic flights
- Flights tracked: VS 25, 26, 3, 4, 45, 46, 9, 10, 127, 128, 137, 138, 47, 48, 153, 154
- Focused on Flight VS4 (JFK → LHR, 18:00 departure) for its consistency (2012–present)

Captured Variables

- Flight Info: Date, Flight No., AC Reg, Gate, On-Ground Time, ATA
- Passenger Data: Last Pax at Customs; Gender (M, F, C, Infants); Class (J, W, Y, Jumpseat); PADs; TOB (Pax, Inf)
- Operational Events: WCHR count; Bag Delivery times; Cargo load; Timeline events (Door Open → Catering Off)

Data Validation & Cleaning

- Aircraft Registration Check: Verified AC Reg against fleet roster
- Missing/Corrupt Data: Dropped rows missing critical fields (e.g., ATA, AC Reg)
- Standardization: Unified time formats, gender/class labels
- Chronology Checks: Ensured event timestamps followed logical order
- Outlier Detection: Flagged and reviewed extreme values
- Deduplication: Removed duplicate log entries



Merge Integrity:

- After joining Visual Crossing weather data, confirmed all VS4 records remained intact
- Checked for and resolved any unmatched or duplicated rows post-merge

Weather Data Integration

- Pulled temperature, precipitation, wind, etc., via Visual Crossing API
- Used to analyze weather's impact on boarding efficiency and delays

MODEL DEVELOPMENT

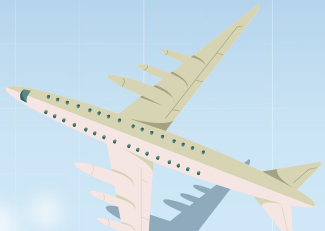
Platform

- Apache Spark 3.4.2 on Hadoop 3

Data File Example

Columns: date, coded_id, month, day, year, day_of_week, registration, seats_j, seats_w, seats_y, temp, humidity, percip, snow, wind_speed, cloud, visibility, final_j, final_w, final_y, final_tob, total_bags, wheelchair, booked_j, booked_w, booked_y, meals_j, meals_w, meals_y, cargo

date	coded_id	month	day	year	day_of_week	registration	seat_j	seat_w	seat_y	temp	humidity	percip	snow	wind_speed	cloud	visibility	final_j	final_w	final_y	final_tob	total_bags	wheelchair	booked_j	booked_w	booked_y	meals_j	meals_w	meals_y	cargo
1/1/2012	1	1	1	2012	1	BLU	45	38	225	46.2	78.2	78.2	0	25.2	45.8	9.8	45	38	223	306	323	3	31	35	228	45	38	225	6,622

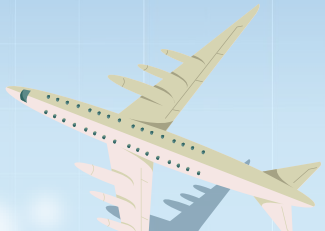


MODEL DEVELOPMENT

Features

- Date: month, day, year, day_of_week
- Configuration: seats_j, seats_w, seats_y
- Weather: temp, humidity, percip, snow, wind_speed, cloud, visibility
- Numbers given T-6hr: #_of_wheelchairs, booked_j, booked_w, booked_y

```
# Define features
feature cols = ["month", "day", "year", "day of week",
                "seats j", "seats w", "seats y", "temp",
                "humidity", "percip", "snow", "wind speed",
                "cloud", "visibility", "wheelchair",
                "booked_j", "booked_w", "booked_y"
]
```



MODEL DEVELOPMENT

Pipeline

```
indexer = StringIndexer(inputCol="registration", outputCol="registrationIndex", handleInvalid="skip")
```

StringIndexer: Converts the string registration → numeric registrationIndex, handleInvalid="skip" drops any unseen or null values during transform

```
imputer = Imputer(inputCols=feature_cols, outputCols=feature_cols)
```

Imputer: Automatically fills nulls in each feature column, Uses column-wise mean by default

```
assembler = VectorAssembler(inputCols=feature_cols + ["registrationIndex"], outputCol="features")
```

VectorAssembler: Concatenates all numeric features + registrationIndex
Produces a single "features" vector column

```
scaler = StandardScaler(inputCol="features", outputCol="scaledFeatures", withStd=True, withMean=True)
```

StandardScaler: Centers data to zero mean & scales to unit variance
Outputs "scaledFeatures" for robust training

```
gbt = GBTRegressor(featuresCol="scaledFeatures", labelCol=label_col, maxIter=100)
```

GBTRegressor: Trains a Gradient-Boosted Trees model on "scaledFeatures"
labelCol=label_col lets us insert final_j, final_w, or final_y
maxIter=100 sets the number of boosting iterations

```
Pipeline(stages=[indexer, imputer, assembler, scaler, gbt])
```

EVALUATION

Train Test Split:
80 Train / 20 Test | (Seed 42)

```
train_data, test_data = flight_df.randomSplit([0.8, 0.2], seed=42)
```

Evaluation Methods:

- *RMSE* (Root Mean Squared Error):
  - RMSE is in the same units as the target (Value of 5 == Off by +/- 5 People)
  - Penalizes big misses because squaring gives extra weight to large errors
- *R^2* (Coefficient of Determination):
  - R^2 tells you how much of the booking-volume mystery the model has unraveled
  - A R^2 score of 0.8 means the model solved 80% of the swings, while only 20% is left as random noise / missing factors.

```
=== final_y ===
RMSE = 0.93, R² = 0.931
Top 10 features:
  booked_y          0.6777
  booked_j          0.0388
  booked_w          0.0326
  temp              0.0306
  day               0.0298
  cloud             0.0286
  humidity          0.0243
  month             0.0239
  wind_speed        0.0202
  day_of_week       0.0191
```

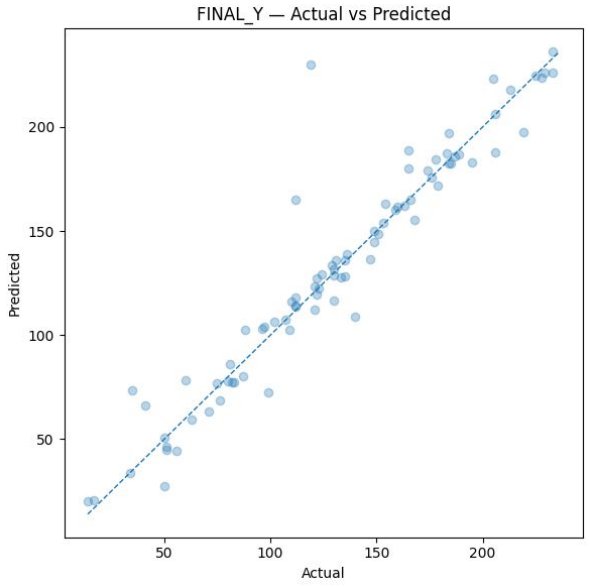
```
=== final_w ===
RMSE = 0.88, R² = 0.879
Top 10 features:
  booked_w          0.5188
  year              0.0668
  booked_j          0.0540
  booked_y          0.0495
  temp              0.0346
  cloud             0.0342
  day               0.0314
  month             0.0313
  humidit           0.0278
registrationIndex0.0246
```

```
=== final_j ===
RMSE = 3.12, R² = 0.897
Top 10 features:
  booked_j          0.5313
  year              0.0502
  month             0.0466
  booked_w          0.0447
  booked_y          0.0410
  temp              0.0376
  seats_j           0.0334
  day               0.0277
  wind_speed        0.0275
  humidity          0.0272
```

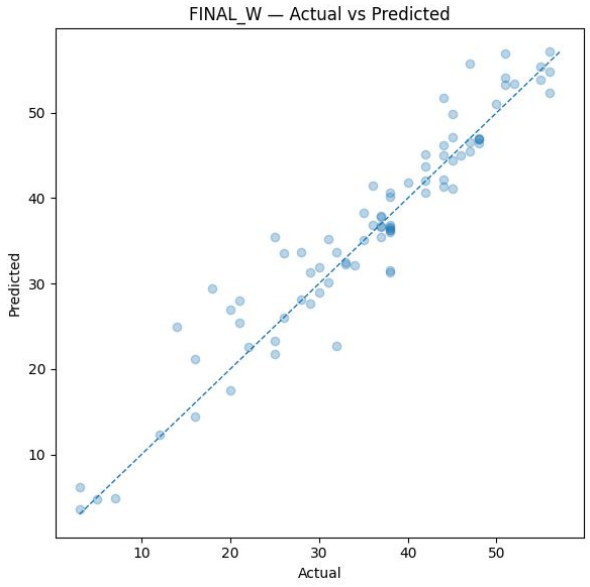
EVALUATION

Actual vs Predicted scatter plots (10% sample):

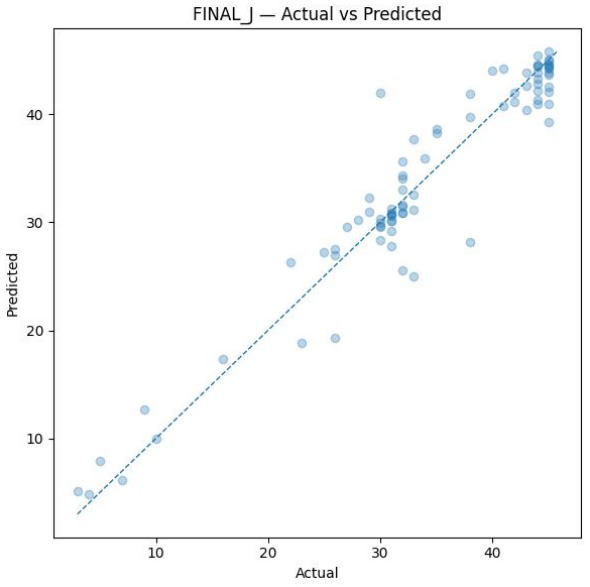
Economy



Premium



Upper Class



PUT INTO PRACTICE

Average Meal Improvement (with Buffer)

80%

Economy

Meals Wasted
- Manually: 8.86 meals
- Prediction: 1.84 meals

~80% improvement

RMSE = 0.93
2 Meal Buffer

18%

Premium

Meals Wasted
- Manually: 2.74 meals
- Prediction: 2.24 meals

~18% improvement

RMSE = 0.88
2 Meal Buffer

-113%

Upper Class

Meals Wasted
- Manually: 1.36 meals
- Prediction: 2.9 meals

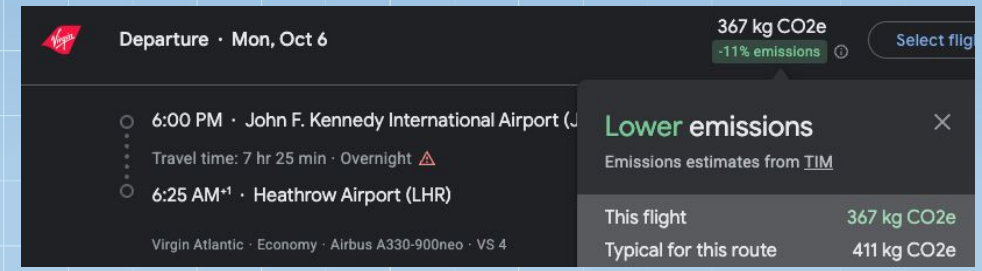
~213% increase in waste

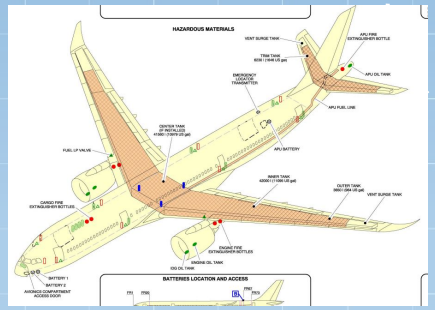
RMSE = 3.12
3 Meal Buffer

PUT INTO PRACTICE

Average Cargo Improvement

- On average there is 10,173 kg of cargo transported
- There is on average ~3x the amount (31,207 kg) of cargo space available
- Allocate 3x amount of cargo to decrease Well-to-Wake emissions in kg CO2e per passenger





On Average **13% Improvement** if 3x cargo is loaded
Estimated emissions in kg CO2e per passenger

Economy	Premium	Upper
343.8 → 299.1	945.7 → 822.8	1719 → 1495.5

Figures for May on a A330-900neo (Calculated through Google's TIM)

WHY DON'T AIRLINES DO THIS YET?

PRIORITIES

The passengers and their bags are the #1 priority.

CONTRACTS

Many clients prefer big freight airlines (FedEx, DHL, Atlas)

UNPREDICTABILITY

Aviation industry is very unpredictable, and things out of your control (weather/OA) can impact operation.

ACCURACY

The model is not 100% accurate and might need help from human Ops Agent / TCO.

OPERATION RISK

Airlines need consistency and compliance and this can introduce risk.

INFRASTRUCTURE

Incorporating this tool in existing legacy systems will be expensive and labor sum.

NEXT STEPS

INTEGRATE REAL-TIME BOOKING FEEDS

Connect the model to the booking system to get real time numbers and have support for dynamic updates.

EXPAND MODEL TO OTHER FLIGHTS

We can expand the model to the other 6 flights we have at virgin and potentially our Manchester flight.

TEAM MEMBER RESPONSIBILITIES

BENOY THOMAS

Benoy acted as a floater looking over the models and checking to see if it met standards or goals our group set for the model itself.

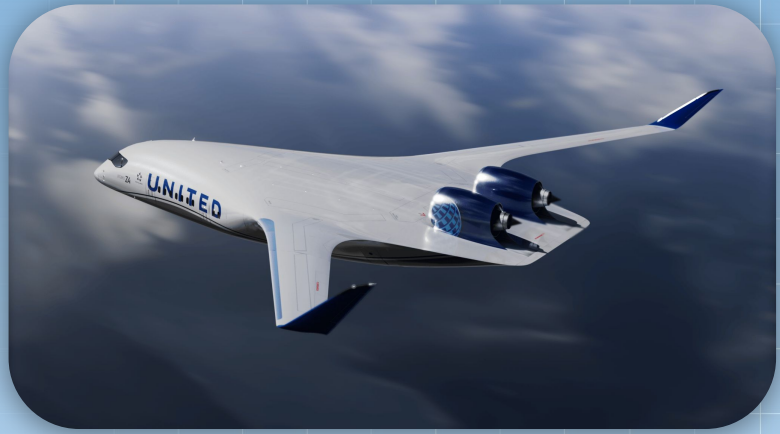
NAVRAJ SINGH

Navraj worked with the dataset and was able to extract information from Virgin Atlantic's database to create our model, specifically V1/3

LUIS JACO

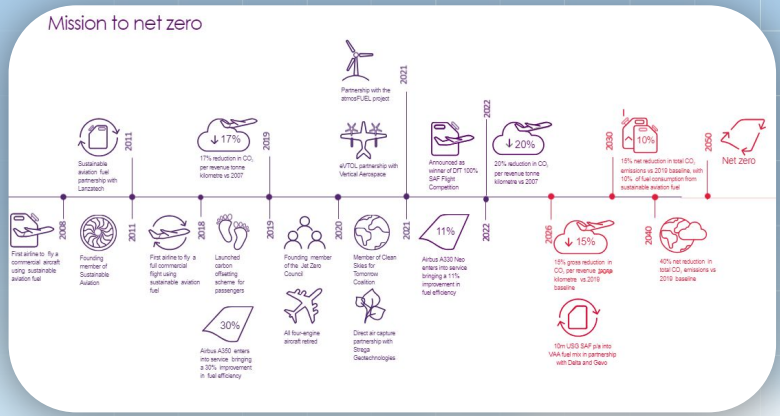
Luis worked on the machine learning model as well specifically V2 in hopes of making the model more accurate for everyday use.

CONCLUSION



HOW IT BENEFITS SOCIETY?

- Makes job easier for all those involved (Ops, TCO, Cargo loaders)
- Reduces meal wastage
- Decreases carbon footprint
- Right now doesn't replace Ops Agent but a tool
- Airlines can invest improving legacy systems rather than create new aircrafts



THANK YOU

DO YOU HAVE ANY QUESTIONS?