IBM Data Science Capstone Final Presentation

Slide 1: Title Slide

Project Title: SpaceX Launch Data Analysis & Prediction\ **Presented by:** Syed Faizan Ali\ **Date:** 7/29/2025\ **Course:** IBM Data Science Capstone

---

Slide 2: Executive Summary

- Collected and cleaned SpaceX launch data from API and Wikipedia
- Performed EDA, SQL analysis, geospatial mapping, and predictive modeling
- Created an interactive dashboard using Plotly Dash
- Key Insight: Payload and launch site influence success rates

---

Slide 3: Introduction

- **Objective:** Analyze SpaceX launches and predict success
- **Why:** Support mission planning and reliability improvements
- **Tools Used:** Python, Pandas, SQL, Folium, Dash, Plotly, Scikit-learn

---

Slide 4: Data Collection

- REST API: SpaceX Launches (v4 endpoint)
- Web Scraping: Wikipedia Falcon 9 page
- Local files for dashboard (CSV)
- SQLite database for structured queries

Slide 5: Data Wrangling

- JSON flattened using `pd.json_normalize`
- Mapped rocket IDs to rocket names
- Filled missing values, formatted date columns
- Loaded into SQLite table: `SPACEXTBL`

Slide 6: EDA & Visual Analytics Methodology

- Used Seaborn & Matplotlib for static visuals
- Plotly Express for interactive graphs
- Grouped by site, outcome, and payload range

Slide 7: SQL Methodology

- Connected via `%sql` magic in Jupyter
- Queried launches per site, payload stats, mission outcomes
- Used subqueries and aggregations

Slide 8: Pie Chart - Successful Launches by Site

Graph: Pie chart showing proportion of successful launches across all sites

---

Slide 9: Scatter Plot - Payload vs Success

Graph: Payload Mass vs Outcome, color-coded by booster version

- Insight: Mid-range payloads had highest success

Slide 10: Success Over Time

Graph: Line chart showing increase in success rate over years

- Insight: Launches became more reliable post-2017

Slide 11: SQL Results - Total Launches per Site

Query:

```
SELECT Launch_Site, COUNT(*) FROM SPACEXTBL GROUP BY Launch_Site;
```

Output Table:

Launch Site	Total Launches
CCAFS SLC 40	55
KSC LC 39A	30
VAFB SLC 4E	15

Slide 12: SQL Results - Min/Max Payload

Query:

```
SELECT MIN(payload_mass__kg_), MAX(payload_mass__kg_) FROM SPACEXTBL;
```

Result:

- Min: 0 kg
- Max: 9600 kg

Slide 13: SQL Results - Mission Outcomes

Query:

```
SELECT Launch_Site, Mission_Outcome, COUNT(*) FROM SPACEXTBL GROUP BY
Launch_Site, Mission_Outcome;
```

Insight: KSC LC 39A has highest success ratio

Slide 14: Folium Map - Launch Sites

- Used `folium.Marker()` and `MarkerCluster`
- Marked: CCAFS, KSC LC-39A, VAFB SLC 4E **Screenshot:** Map image with all sites

---

Slide 15: Folium Map - Success/Failure Markers

- Colored markers for outcome (green = success, red = failure)
- Distance to coast/city calculated using Haversine formula

Slide 16: Dash Dashboard - Pie Chart

Screenshot: App section showing dropdown and pie chart for site selection

---

Slide 17: Dash Dashboard - Scatter Plot

Screenshot: Scatter showing Payload vs Outcome filtered by site & slider

---

Slide 18: ML Models & Evaluation

- Models: Logistic Regression, SVM, Decision Tree
- Best Accuracy: **83.33%** using Decision Tree
- Features used: Payload, Site, Booster Version

Slide 19: Confusion Matrix

Chart: Visualizing true positives, false positives, etc.

- Model validated with holdout test set (18 samples)

Slide 20: Conclusion

- High payload reduces success probability
- Most reliable site: KSC LC-39A
- Reusable boosters improve outcomes
- Dashboard & maps offer intuitive insight

Slide 21: Creativity / Insights

- Correlated booster version with reusability and success
- Filtered data using widgets
- Map animations (optional GIF)

Slide 22: Thank You!

GitHub Repo: https://github.com/SyedFaizanAlii\ **Questions?**