

# Manifesto: Embracing Data

By Beatrice Archer

## I. Understanding Data

Data isn't just a collection of numbers or facts; it's the fabric upon which our understanding of the world is woven. It encompasses everything from the mundane to the extraordinary—the leaves scattered on the ground, the patterns of our sleep, the breaking of a routine schedule. It's both omnipresent and ephemeral. What makes it 'data' is our choice to perceive, track, and comprehend it.

The evolution of humanity's perception of data is remarkable. From a negligible afterthought to an omnipresent force in our lives, data has exploded in significance over centuries. Its transformation from a passive existence to an active entity reflects our changing relationship with information. It's not merely a product of our technological advancements; it's a product of our curiosity and desire to understand the world around us.

## II. The Role of a Data Scientist

A data scientist isn't merely defined by technical skills; it's a versatile mindset, a way of thinking. It involves navigating data's complexities, from collection to transformation into insights. Critical thinking is crucial, but adaptability is our true asset.

The nature of data science projects is inherently dynamic. Data science projects aren't static; they evolve, demanding flexibility and the ability to pivot as insights unfold and challenges arise. An illuminating example emerged during a web scraping project where the exploration of various websites to extract desired data not only showcased technical learning but emphasized persistence and adaptability. I explored about ten different sites that I couldn't locate the div for collecting the data I wanted. This revealed the importance of perseverance amid learning curves and project ideation.

## III. Advice for Aspiring Data Scientists

Find enjoyment in the process of data exploration. It's not merely about manipulating numbers but discovering patterns and telling stories. Flexibility is key; be open to unforeseen discoveries that might reshape your project's direction.

Flexibility is an essential trait. Embracing unexpected turns; the best discoveries often emerge from uncharted paths. Reflecting on my airport data project, initially conceived as a heat map visualization, the journey evolved into a visually captivating rotating globe. This shift represented the joy and creativity found in exploring unforeseen opportunities within data science. This is the link to the project feel free to check it out: <https://observablehq.com/@beabuz-ws/international-airports-visulazation>

## IV. Problem-solving with Data Analysis and Visualization

Data analysis and visualization hold immense power. They provide us with the tools to unravel mysteries and answer questions across a vast spectrum of topics. However, understanding data's limitations is crucial. Some questions remain unanswered due to data constraints or subjectivity. Midway through

projects, we encountered issues with datasets lacking essential information or poorly formatted data, prompting adaptations in our approach. You learn to compromise and change either the questions you're asking, or the data set itself. In the airport data project, discrepancies in flight data prompted a shift in focus towards location-based insights. If you look at the image below you can see the airports stored in order of number of flights.

```
# This counts departures from each airport
departures = flight['usg_apt'].value_counts().reset_index()
departures.columns = ['Airport', 'DepartureCount']

# This counts arrivals at each airport
arrivals = flight['fg_apt'].value_counts().reset_index()
arrivals.columns = ['Airport', 'ArrivalCount']

# This merges the counts from departure and arrival into a single DataFrame
total_flights = pd.merge(departures, arrivals, on='Airport', how='outer').fillna(0)
total_flights['TotalFlights'] = total_flights['DepartureCount'] + total_flights['ArrivalCount']

# Airport with the most flights
busiest_airport = total_flights[total_flights['TotalFlights'] == total_flights['TotalFlights'].max()]

print(total_flights) # Displays the total flights for each airport
print("Busiest Airport:")
print(busiest_airport) # Display the busiest airport
```

	Airport	DepartureCount	ArrivalCount	TotalFlights
0	JFK	64122.0	0.0	64122.0
1	MIA	61939.0	0.0	61939.0
2	LAX	47760.0	0.0	47760.0
3	ORD	34952.0	0.0	34952.0
4	EWB	33024.0	0.0	33024.0
...	...	...	...	...
2135	TYL	0.0	1.0	1.0
2136	SXU	0.0	1.0	1.0
2137	RKT	0.0	1.0	1.0
2138	PUG	0.0	1.0	1.0
2139	SHB	0.0	1.0	1.0

[2140 rows x 4 columns]  
Busiest Airport:

	Airport	DepartureCount	ArrivalCount	TotalFlights
0	JFK	64122.0	0.0	64122.0

You can see JFK is the busiest airport with 64122 departures. However, the strange thing is that it had no arrivals how the busiest airports can have no arrivals. That's when we realized that our data was not a full comprehensive list of flights like it claimed it was weirdly missing parts. So, we simply had to do the projects focused on the airport's location and not the flights. Luckily it worked out because we could still do the globe with the airport location, and we simply focused on understanding the data set we had for flights and presenting what it had.

## V. Principles of Data Science Process

Flexibility: Embrace project evolution and adapt as insights emerge.

Enjoyment: Find fulfillment in uncovering patterns and stories within the data.

Data's Limits: Recognize constraints; not all questions have clear answers. You may need to change your data set.

Subjectivity: Acknowledge that interpretations may vary based on context and perspective. If you have a bias data set, make that clear to the viewer.

#### VI. Application and Communication

Apply these principles to new projects, embracing adaptability and creativity. Communicate the essence of data science through practical examples and discussions, making its core tenets accessible to newcomers.

In summary, data isn't just numbers; it's a dynamic force shaping our world. By understanding its complexities and limitations, and finding joy in the process, we transform raw data into meaningful narratives.