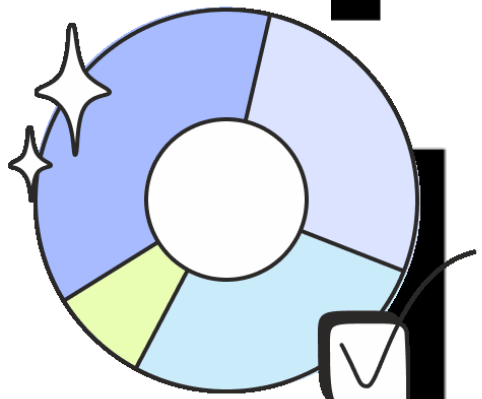
Agenda

1. Business Case & Objectives
2. Methodology Overview
3. Strategy Integration
4. Technology Requirements
5. Project Management Approach
6. Team Structure
7. Ethical, Privacy, and Legal Considerations

1 Business Case & Objectives



ASOS facing historic performance challenges

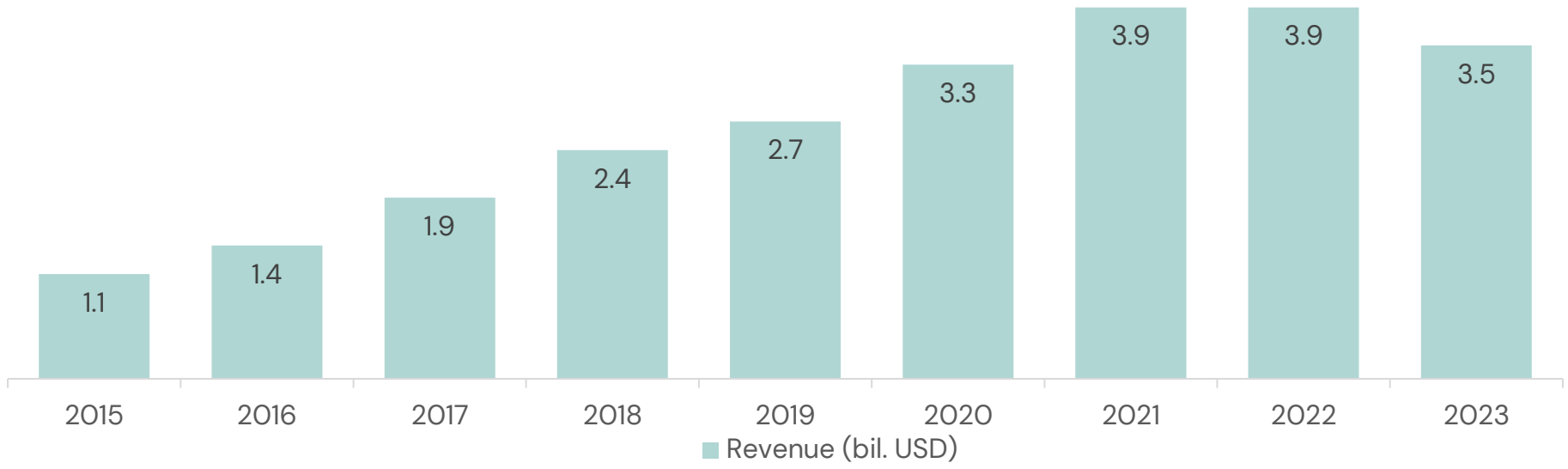
Asos falls £291m into the red as shoppers return to high street

'Challenging trading backdrop' for online fashion retailer, with sales down more than expected

Asos expects sales slump to continue as it is hit by near £300m loss

Shares plunge after analysts say online fashion retailer may need to raise more cash, possibly through sale of Topshop brand

ASOS' revenue throughout the years



Huge loss from increasing customer returns

Asos warns on profits amid 'significant increase' in customer returns

Sharp share sell-off as retailer says inflation is giving customers second thoughts about purchases

Fast fashion retailer ASOS struggles to engage customers as losses balloon

One expert told The Independent that its free returns model makes the company susceptible to economic instability

Some experts now say that the holes within ASOS's business model are becoming apparent. Rick Smith, Managing Director of business recovery firm Forbes Burton tells *The Independent* that ASOS's business model of allowing large numbers of free returns makes the company susceptible to financial losses during challenging economic conditions.

Many unhappy returns: UK retailers count the costly growth in sent-back items

Mounting waste, rising costs and shrinking profits prompt firms to find ways to deter shoppers from sending stuff back

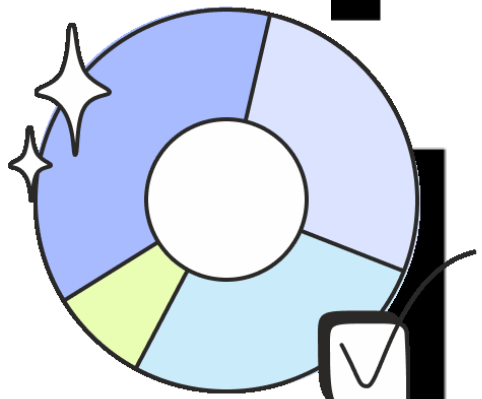
Boohoo swings to £91m loss as shoppers return more items

Online fashion business loses out as shoppers return to high streets after Covid lockdowns

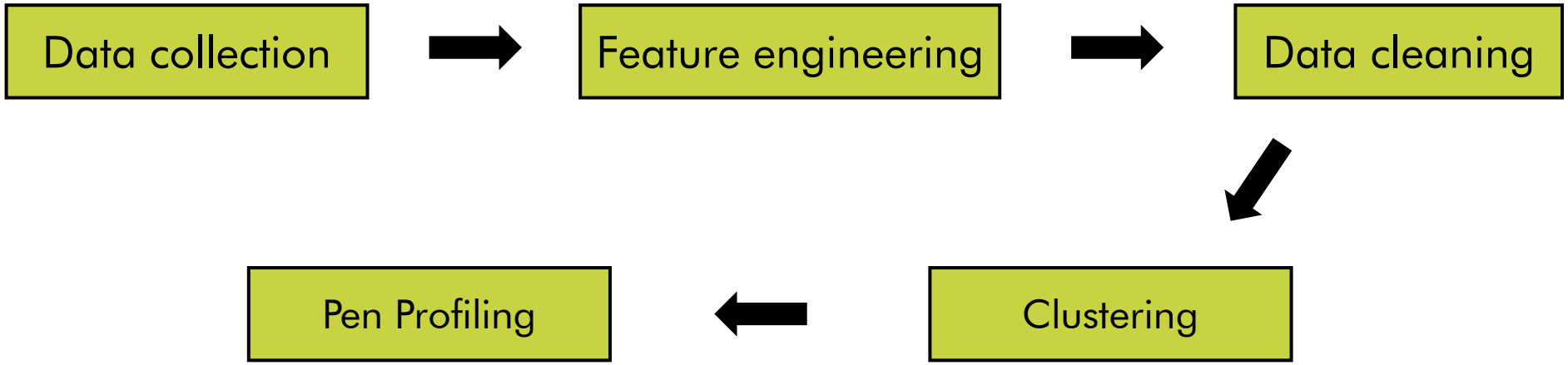
Our solutions

- Employ sophisticated and data-driven **customer segmentation methods** to group similar customers in terms of characteristics and return behavior.
- Develop **highly targeted returns policies and marketing strategies** for each customer segment.
- Identify and address **areas of potential loss caused by returns**, implementing proactive measures to safeguard profitability.

2 Methodology Overview



Methodology overview

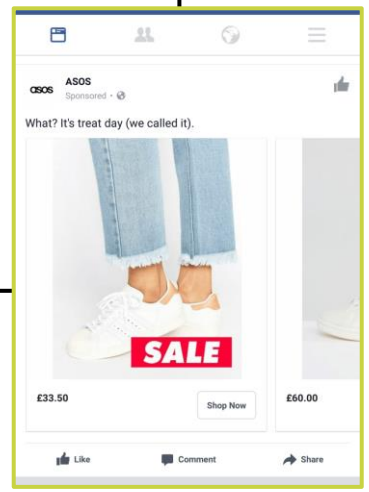


Methodology overview | Data collection

During its operation, ASOS deals with an extremely high volume, variety, and velocity of data generated from various media sources. Have you made use of all of this data?

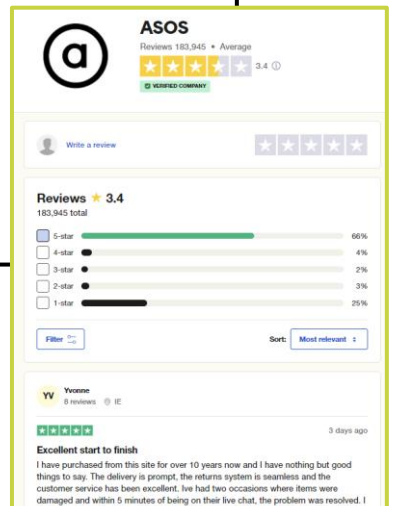
Paid media

Advertising:
Click-through rates, Conversion rates



Earned media

Review Sites:
Reviews, Ratings, Product satisfaction, Preferred products



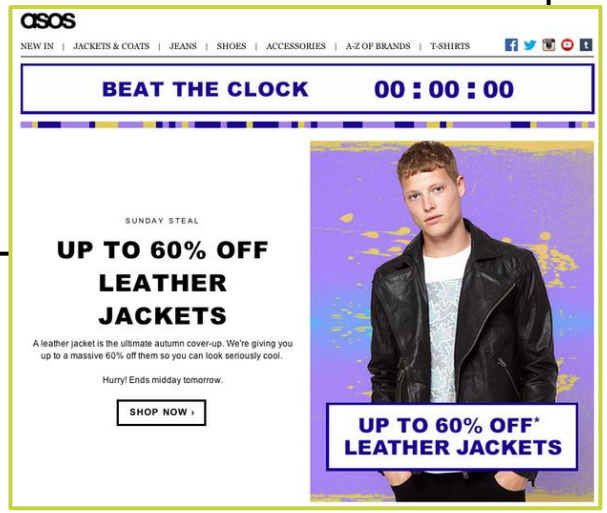
Owned media

Website, Mobile App: Preferences; Wishlist items; Browsing history; Clickstream data; Items viewed, added to cart, and purchased; Purchase behavior

Loyalty Program: User account information; Purchase history; Preferred products and categories

Email Campaigns: Open rates, Click-through rates, Conversion rates, Customer engagement, Preferred products

Virtual Try-on Features: Products tried on, Time spent using the feature, Style preferences



Social media

Social Media: Likes, shares, comments, Mentions of ASOS

We will leverage data from various media sources for our project, ensuring a comprehensive approach to data utilisation.

Methodology overview | Data collection

"For businesses to form a complete picture, they need both big and thick data because each of them produce different types of insights at varying scales and depths." (Wang, 2016)

We'll integrate **"thick data"** derived from **bespoke data collection** for a more **comprehensive and in-depth data analysis**:

- A survey of 10,000 loyal customers will be conducted to collect: demographic data (e.g., gender, age, location), psychographic data (e.g., personality type, interests, values), behavioral data (e.g., categories of items purchased, brands, frequency, …).
- Relevant features from the survey data will be extracted and matched with corresponding clusters to enhance the understanding of each segment.

Methodology overview | Feature engineering

Unique Identifier	
customer_id	Unique identifier for each customer
Demographics	
age	Age of the customer
Purchase behavior	
total_transaction	Total number of transactions made by the customer.
total_spend	Total amount spent by the customer on purchases
days_since_last_purchase	Number of days since the customer's last purchase
average_order_value	Average value of orders placed by the customer
average_item_purchased	Average number of items purchased per transaction
avg_item_spend	Average spending per item
total_categories_purchased	Total number of unique product categories purchased by the customer
Website interactions	
past_visits	Number of past visits to the website or app
total_pages_viewed	Total number of pages viewed by the customer during their visits
time_per_page	Average time spent on each page
avg_visit_duration	Average duration of each visit
campaign_click_through	Click-through rates on email campaigns

Methodology overview | Feature engineering

Return history	
return_frequency	Total number of returns by the customer
total_spend_on_returns	Total amount spent by the customer on returned items
total_categories_returned	Total number of unique product categories returned by the customer
top_returned_categories	Most frequently returned product categories by the customer
return_rate	The percentage of products returned by customers compared to the total number of products purchased
Sentiments	
avg_sentiment_score	Score indicating the sentiment (positive, negative, neutral) expressed in the customer review.

Methodology overview | Clustering

Dealing with big data: 23.3m active customers (2023)

BIRCH (Zhang, Ramakrishnan, & Livny, 1997)

How BIRCH works?

- Data is summarized into compact Clustering Features (CF) entries.
- These entries are organized in a CF tree, a height-balanced tree with each leaf node containing a summary of CF entries.
- BIRCH scans the data, inserts points into the tree, and adjusts the threshold value if it runs out of memory.
- It groups crowded subclusters, applies a clustering algorithm to categorize CFs

Why BIRCH?

- **Scalability**: Due to its incremental approach and compact data representation.
- **Efficiency**: With only a single scan of the dataset, BIRCH can produce a good clustering result. No need to define the number of clusters
- **Quality of Clusters**: Even when clusters are of arbitrary shape.
- **Robustness to Outliers**

Methodology overview | Pen Profiling

After forming clusters, each is individually analysed using numerical data fed into the clustering algorithm, along with available categorical data and survey data, to define its unique characteristics. They are then assigned a name and description. Below is an example of potential segments resulting from the algorithm's execution:

Young Urbanites

Bulk Buyers

Luxury Lovers

Occasional Shoppers

Impulsive Shoppers

Tech-savvy Customers

Loyal Customers

Methodology overview | Pen Profiling

Young Urbanites

- Young urban dwellers
- Regular clothing and shoe buyers, active online
- Frequent returns due to sizing concerns

Bulk Buyers

- Influencers, stylists, or businesses
- High total transaction counts and spend
- High return rates due to large-scale styling or inventory needs.

Luxury Lovers

- Higher income groups
- Prefer premium brands, making fewer but larger transactions
- Rarely return, typically due to quality concerns

Impulsive Shoppers

- Quick browsing, high click-through rates
- Frequent returns due to regret or dissatisfaction with impulse buys.

Occasional Shoppers

- Newcomers to the brand
- Limited purchase history, sporadic browsing. Increased activity during holidays, engaging more with promotions
- Moderate return rates, primarily for fit or quality issues

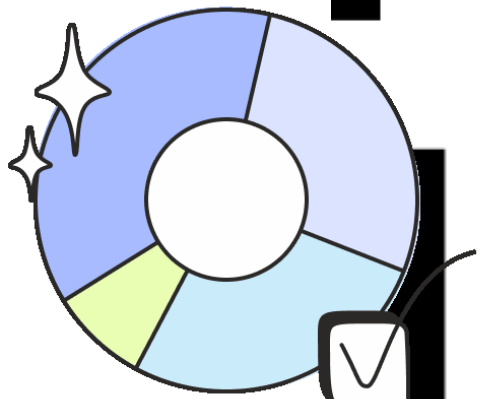
Tech-savvy Customers

- Heavy engagement with the brand's digital aspects
- High browsing activity and interactions with digital touchpoints
- Lower return rates due to more informed purchasing decisions

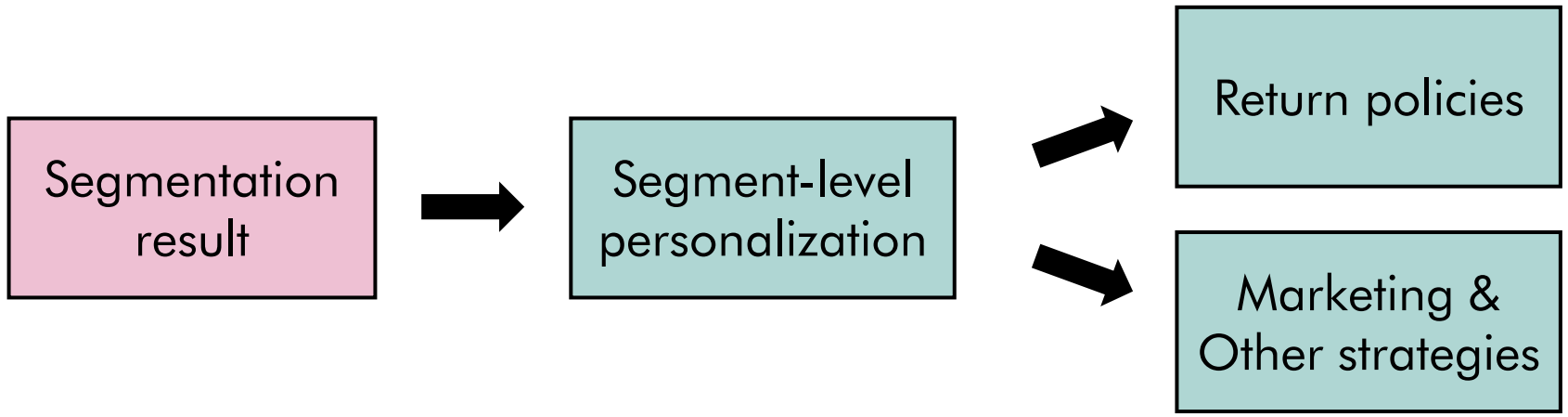
Loyal Customers

- Regular purchases, high total transaction counts and spend. Positive brand interactions: frequent website visits and engagement with email campaigns
- Rarely return items due to familiarity with the brand

3 Strategy Integration



'Companies that excel at personalization generate 40% more revenue from those activities than average players' (McKinsey, 2021)



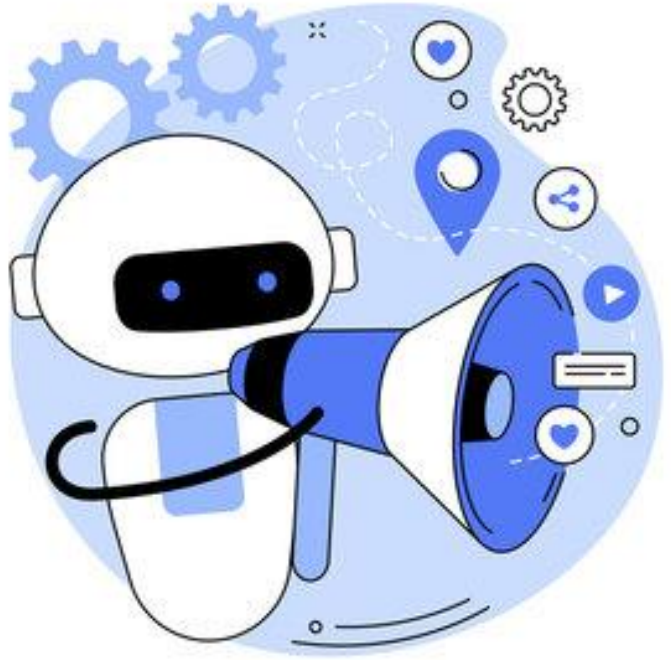
Personalised return policies

An example of personalised return policies for potential customer segments. Return policies are communicated clearly at time of purchase and regularly reviewed based on return rates and feedback.

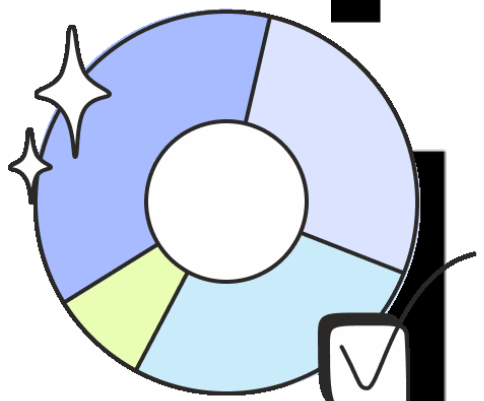
Young Urbanites	Extend the return window
Bulk Buyers	Tiered return policy: Free returns for up to 20% of the order Credit for future purchases instead of refunds Pre-return consultation to try and resolve the issue
Luxury Lovers	Maintain the standard return window, but offer premium return services: prepaid labels, return pickups
Impulsive Shoppers	Instant return option within 2 hours of purchase, after which standard return window applies. Shorter return windows for seasonal items
Occasional Shoppers	Shorter return window with the option for store credit instead of refunds.
Tech-savvy Customers	Extend return windows Tech troubleshoot before returns
Loyal Customers	No change

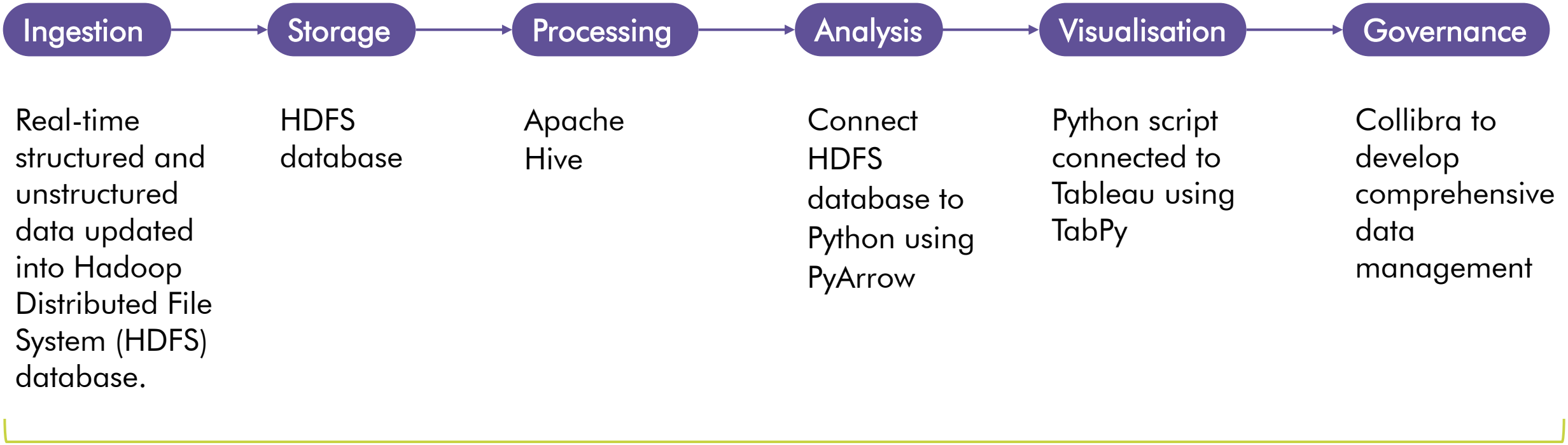
Personalised marketing and other strategies

- Recommender System that suggests products with a lower return rate within each segment
- Personalised shopping consultations (via chat or video) for high-value items to discuss product details, quality, and fit before purchase
- Send reminders based on the customer's size and show choices made by others with the same size if they are picking a different size.
- Provide guides tailored to first-time buyers to help them navigate the purchasing process.

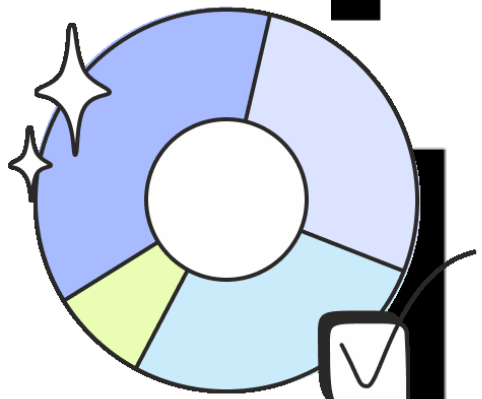


4 Technology Requirements




Ingestion
Storage
Processing
Analysis
Visualisation
Governance
Real-time structured and unstructured data updated into Hadoop Distributed File System (HDFS) database.
HDFS database
Apache Hive
Connect HDFS database to Python using PyArrow
Python script connected to Tableau using TabPy
Collibra to develop comprehensive data management
Apache Airflow to implement robust workflow orchestration and scheduling mechanisms

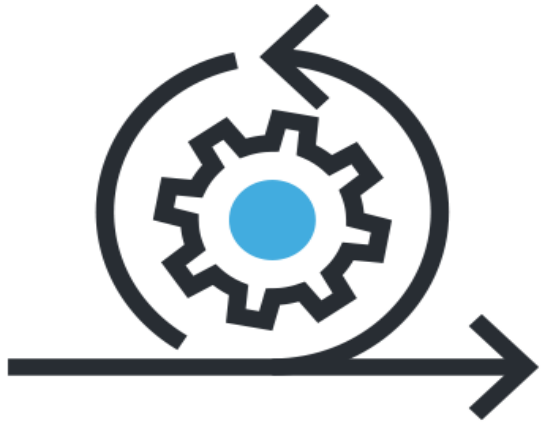
5 Project Management



Agile methodology for efficient project delivery

Sprint-powered momentum: Projects broken down into manageable sprints for tangible progress.

Reflective iteration: Pause, reflect, and refine after each sprint



Accelerated progress & continuous improvement: A culture of ongoing enhancement to drive project excellence.

Customer-centric approach: Align our every step with customer needs

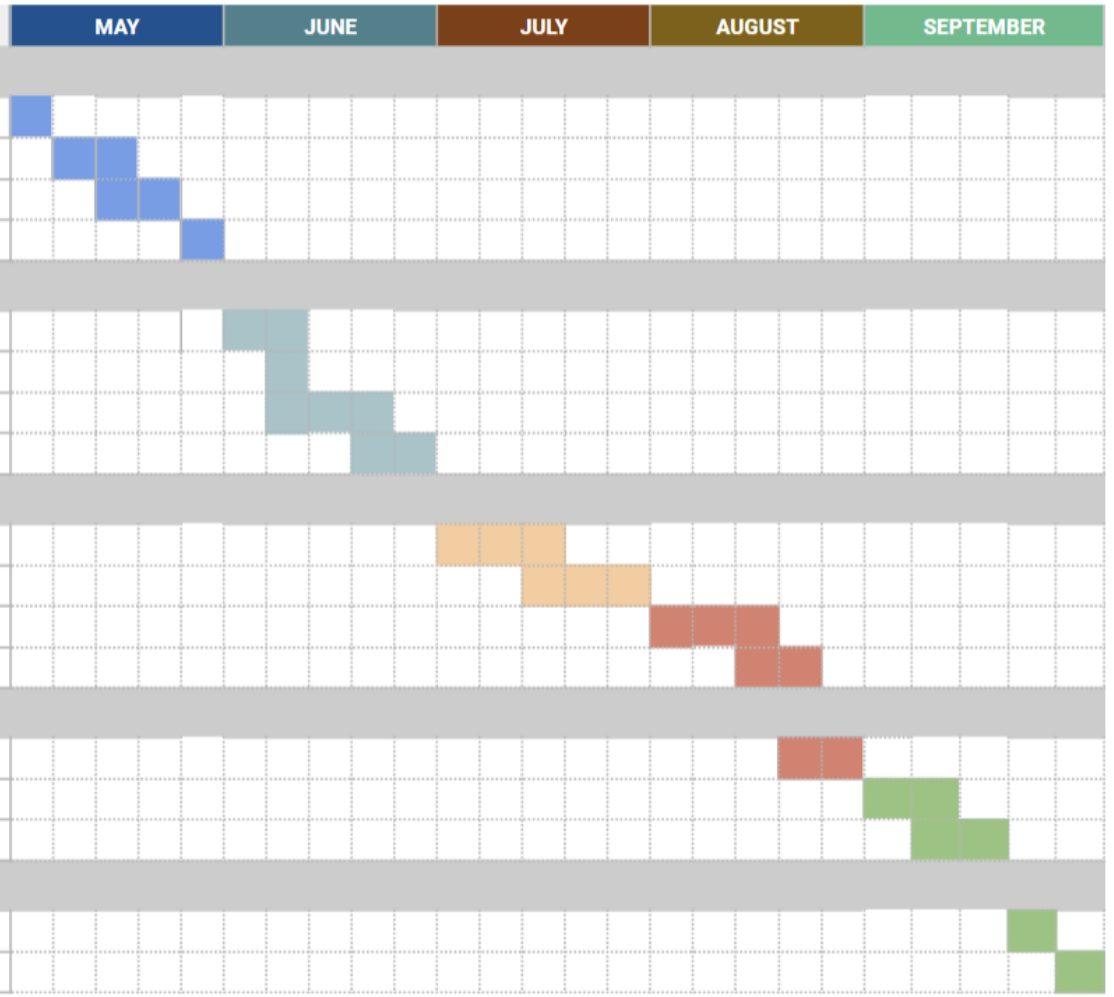
Estimated project plan

PROJECT TITLE: Leading Big Data Business Project

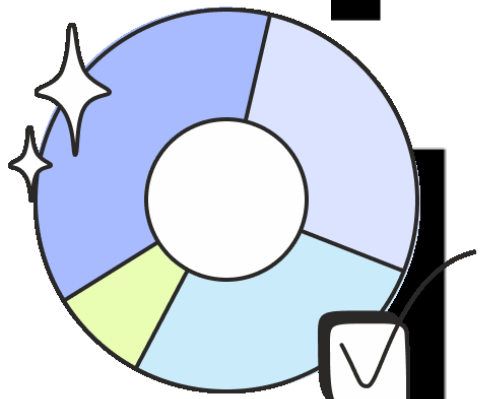
COMPANY NAME: DataPulse Innovations

DATA PULSE INNOVATIONS

SERIAL NUMBER	TASK TITLE	START DATE	DUE DATE	DURATION
1	**Stakeholder Requirements**			
1.1	Meetings with Stakeholders	13/05/2024	14/05/2024	2
1.1.1	Website/App review	15/05/2024	17/05/2024	3
1.2	Problem Identification	18/05/2024	20/05/2024	3
1.3	Workplan	21/05/2024	22/05/2024	2
2	**Data Collection and Preparation**			
2.1	Data Transformation	23/05/2024	01/06/2024	10
2.2	Data Exploration	02/06/2024	06/06/2024	5
2.3	Data Visualization	07/06/2024	21/06/2024	15
2.4	Feature Selection and Engineering	22/06/2024	01/07/2024	10
3	**Machine Learning Modelling**			
3.1	Model Selection	02/07/2024	16/07/2024	15
3.2	Data Training and Validation	17/07/2024	31/07/2024	15
3.2.1	Data Testing	01/08/2024	15/08/2024	15
3.2.2	Performance Metrics Evaluation	16/08/2024	25/08/2024	10
4	**Delivery**			
4.1	Dashboard Building	26/08/2024	30/08/2024	5
4.2	User Testing	31/08/2024	04/09/2024	5
4.3	Code Documentation	05/09/2024	09/09/2024	5
5	**Reporting**			
5.1	Final Documention	10/09/2024	13/09/2024	3
5.2	Presentation	14/09/2024	16/09/2024	3

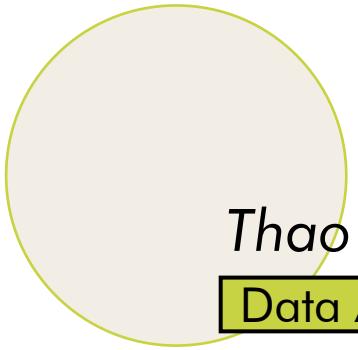


6 Team Structure



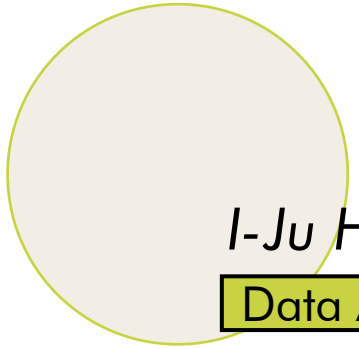
Wania Zaheer

Data Engineer



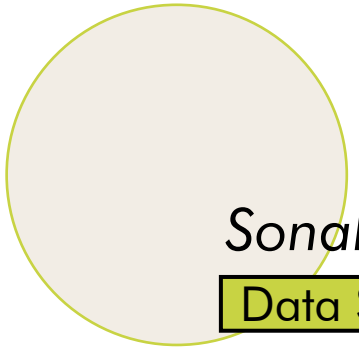
Thao Le

Data Analyst



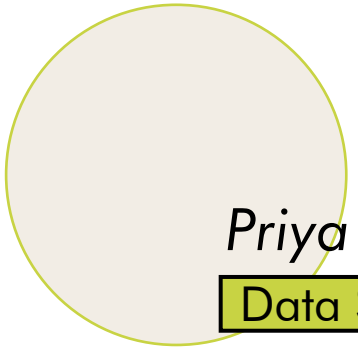
I-Ju Huang

Data Analyst



Sonal Singh

Data Scientist



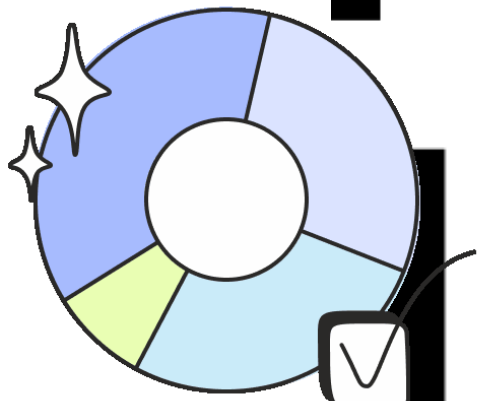
Priya Rathore

Data Scientist

Roshni Namole

Business Analyst

7 Ethical, Privacy, and Legal Considerations



Implications that may arise	How we'll address them
Privacy Concerns: Customers may feel their privacy is compromised if their buying and returning behavior is extensively analyzed (Miyazaki & Fernandez, 2000)	• Transparent communication about data usage: • Opt-out options • Data protection measures: Data minimization, encryption of data, access controls, regular data backups
Data Bias and Discrimination: Unintentional bias in data analysis or policy implementation could lead to discrimination against certain customer segments (Schmitz, 2014)	• Diverse data collection • Algorithmic fairness assessments • Regular audits of policy outcomes • Document the provenance, creation, and use of datasets with 'datasheets for datasets' (Gebru et. al, 2021)
Security Risks: Data breaches could lead to unauthorized access to sensitive customer information.	• Encryption, access controls, regular data backups to safeguard customer data
Legal Compliance: Failure to comply with data protection laws may result in legal repercussions.	Adhere to legal frameworks (e.g., GDPR) by: • Obtaining consent for data processing • Collecting only necessary data for analysis • Ensuring data accuracy and currency • Deleting customer data post-project completion