

# Data Governance @ SneakerPark

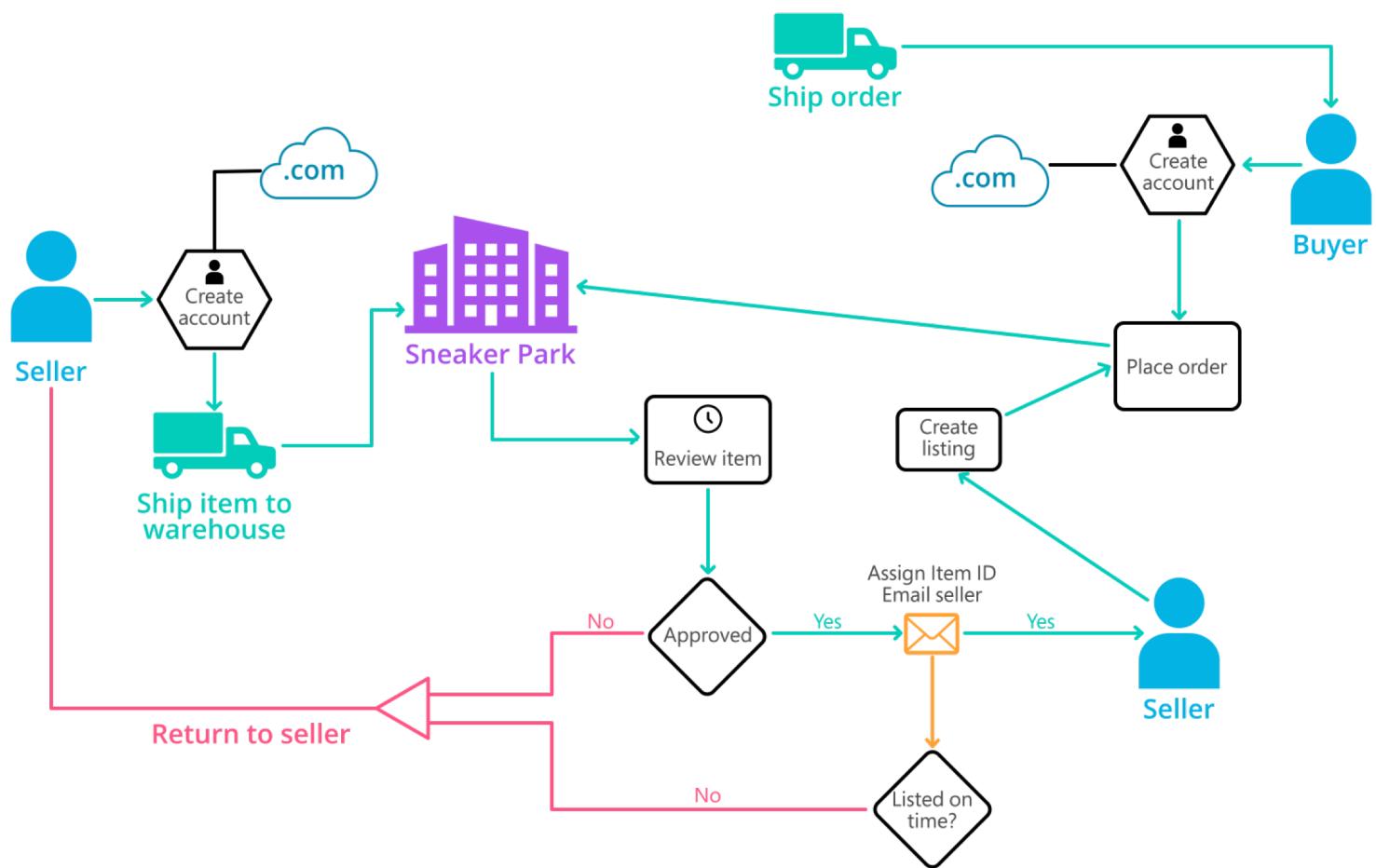


# Background

- **SneakerPark** is an online shoe reseller that allows people to buy and sell used and new shoes. Buyers can bid for shoes or buy them outright, and sellers can set a price or sell to the highest bidder.
- Each buyer and seller must have an active account in order to sell, bid, or purchase sneakers using SneakerPark's website.
- SneakerPark authenticates the shoes before shipping them to the buyer, so before listing an item, the seller must ship it to SneakerPark's warehouse. Upon receipt, SneakerPark assigns an item number to each pair of sneakers and notifies the seller that they are now free to list their item. If the item is not listed within 45 days, SneakerPark returns it to the seller and sends an invoice to the seller for the shipping cost.
- If the item is found to be inauthentic or in an unacceptable condition, it is also returned back to the seller in a similar fashion.
- When the item sells, the buyer's account is credited with the purchase price minus the SneakerPark service fee and shipping fees to deliver the item to the buyer.
- Currently, SneakerPark only supports sales within the United States.

# Background (cont'd)

- Below is a diagram for some of SneakerPark's business processes.



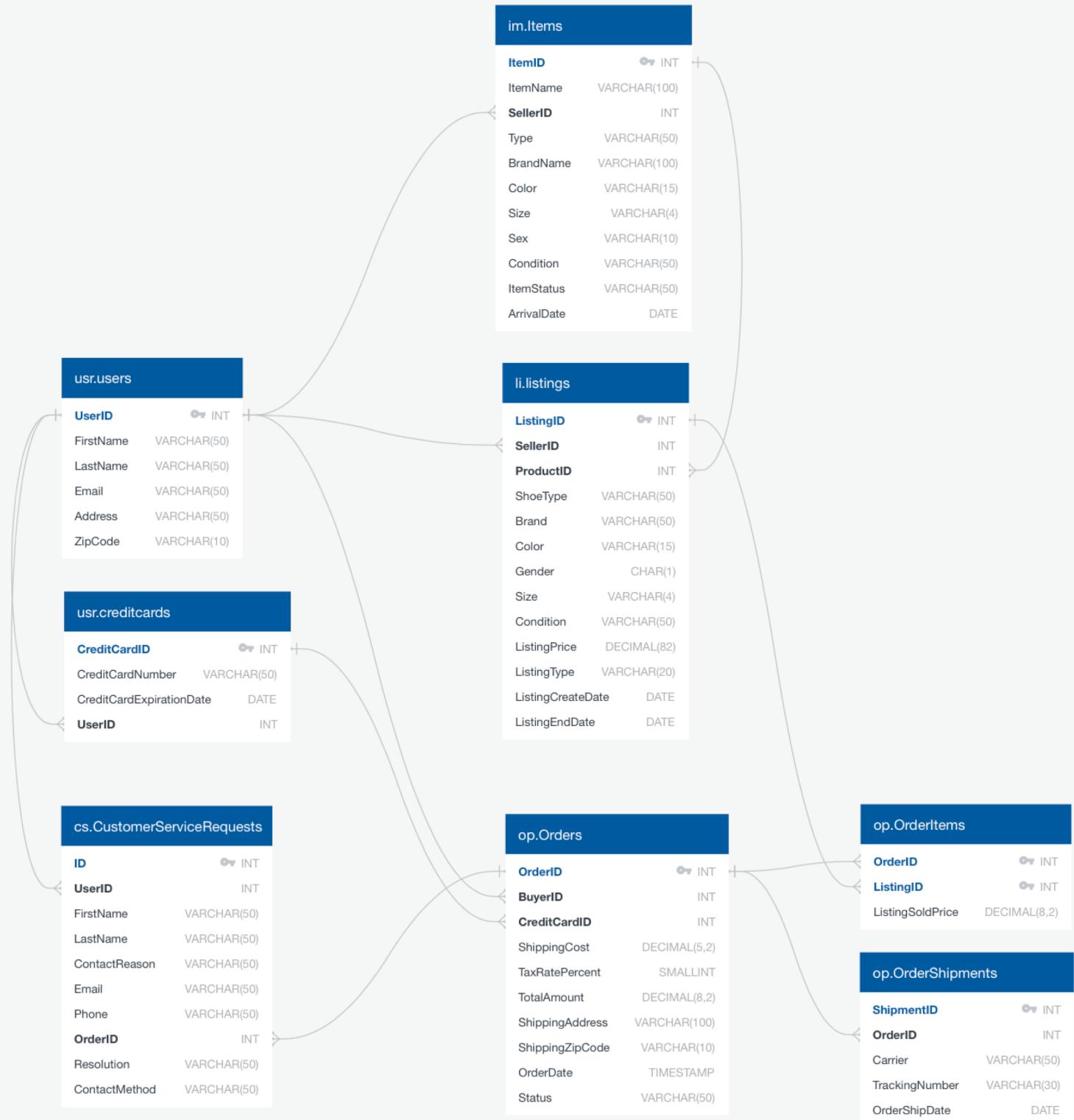
---

## **Step 1**

### Enterprise Data Catalog

#### Part 1: Enterprise Data Model

The Enterprise Data Model provides a holistic view of SneakerPark's data systems and the organization's important entities and relationships.



---

## **Step 2**

# Enterprise Data Catalog

## Part 2: Metadata

# Technical Metadata

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Source System	Schema	Entity	Table Name	Column Name	Description	Data Type	Required	Unique	Value Example Min	Value Example Max	Primary Key	Foreign Key	Foreign Key Table	Foreign Key Column
2	Customers	usr	Users	users	UserID	User ID	INT	Y	Y	121	99900	Y	N		
3	Customers	usr	Users	users	FirstName	First Name	VARCHAR(50)	Y	N	Abel	Zana	N	N		
4	Customers	usr	Users	users	LastName	Last Name	VARCHAR(50)	Y	N	Agudelo	Wire	N	N		
5	Customers	usr	Users	users	Email	Email	VARCHAR(50)	Y	Y	abel.kolstad@fakeemail.com	zana.lehy@infinity.com	N	N		
6	Customers	usr	Users	users	Address	Address	VARCHAR(50)	Y	N	0 Avon Place	9988 Haddow Grove	N	N		
7	Customers	usr	Users	users	ZipCode	Zip Code	VARCHAR(10)	Y	N	1013	98814	N	N		
8	Customers	usr	Creditcards	creditcards	CreditCardID	Credit Card Id	INT	Y	Y	404	99838	Y	N		
9	Customers	usr	Creditcards	creditcards	CreditCardNumber	Credit Card Number	VARCHAR(50)	Y	N	1.22879E+13	9.96585E+13	N	N		
10	Customers	usr	Creditcards	creditcards	CreditCardExpirationDate	Credit Card Expiration Date	DATE	Y	N	44563	46017	N	N		
11	Customers	usr	Creditcards	creditcards	UserID	User Id	INT	Y	N	160	99900	N	Y	users	UserID
12	Orders	op	Orders	Orders	OrderID	Order id	INT	Y	Y	92	31787	Y	N		
13	Orders	op	Orders	Orders	BuyerID	Buyer Id	INT	Y	N	160	99900	N	Y	users	UserID
14	Orders	op	Orders	Orders	CreditCardID	Credit Card Id	INT	Y	N	1499	97365	N	Y	creditcards	CreditCardID
15	Orders	op	Orders	Orders	ShippingCost	Shipping Cost	DECIMAL(52)	Y	N	1.17	19.84	N	N		
16	Orders	op	Orders	Orders	TaxRatePercent	Tax Rate Percent	SMALLINT	Y	N	5	9	N	N		
17	Orders	op	Orders	Orders	TotalAmount	Total Amount	DECIMAL(82)	Y	N	0	237.68	N	N		
18	Orders	op	Orders	Orders	ShippingAddress	Shipping Address	VARCHAR(100)	N	N	0 Courage Drive	9988 Haddow Grove	N	N		
19	Orders	op	Orders	Orders	ShippingZipCode	Shipping Zip Code	VARCHAR(10)	Y	N	2881	98814	N	N		
20	Orders	op	Orders	Orders	OrderDate	Order Date	TIMESTAMP	Y	N	2020-11-01T18:18:12.000000000	2020-12-31T01:41:01.000000000	N	N		
21	Orders	op	Orders	Orders	Status	Status	VARCHAR(50)	N	N	Paid	Shipped	N	N		
22	Orders	op	Order Shipments	OrderShipments	ShipmentID	Shipment Id	INT	Y	Y	50249	940280	Y	N		
23	Orders	op	Order Shipments	OrderShipments	OrderID	Order id	INT	Y	N	427	29692	N	Y	Orders	OrderID
24	Orders	op	Order Shipments	OrderShipments	Carrier	Carrier	VARCHAR(50)	Y	N	DHL	USPS	N	N		
25	Orders	op	Order Shipments	OrderShipments	TrackingNumber	Tracking Number	VARCHAR(30)	N	N	43B4D7B41513M2MI	ZP0L3WZB2DA6M6XK	N	N		
26	Orders	op	Order Shipments	OrderShipments	OrderShipDate	Order Ship Date	DATE	Y	N	44140	44186	N	N		
27	Inventory	im	Items	Items	ItemID	Item Id	INT	Y	Y	509	99996	Y	N		
28	Inventory	im	Items	Items	ItemName	Item Name	VARCHAR(100)	Y	N	Alethea	Ulysses	N	N		
29	Inventory	im	Items	Items	SellerID	Seller Id	INT	Y	N	121	99900	N	N	users	UserID
30	Inventory	im	Items	Items	Type	Type	VARCHAR(50)	Y	N	Boots	Sneakers	N	N		
31	Inventory	im	Items	Items	BrandName	Brand Name	VARCHAR(100)	Y	N	ASICS	Vans	N	N		
32	Inventory	im	Items	Items	Color	Color	VARCHAR(15)	Y	N	black	yellow	N	N		
33	Inventory	im	Items	Items	Size	Size	VARCHAR(4)	Y	N	0	9	N	N		
34	Inventory	im	Items	Items	Sex	Sex	VARCHAR(10)	Y	N	Female	Male	N	N		
35	Inventory	im	Items	Items	Condition	Condition	VARCHAR(50)	Y	N	like new	used	N	N		
36	Inventory	im	Items	Items	ItemStatus	Item Status	VARCHAR(50)	N	N	approved	sold	N	N		
37	Inventory	im	Items	Items	ArrivalDate	Arrival Date	DATE	N	N	44099	44135	N	N		
38	Listings	li	Listings	listings	ListingID	Listing Id	INT	Y	Y	318	999811	Y	N		
39	Listings	li	Listings	listings	SellerID	Seller Id	INT	Y	N	121	99900	N	Y	users	UserID
40	Listings	li	Listings	listings	ProductID	Product Id	INT	Y	N	509	99838	N	Y	items	ItemID
41	Listings	li	Listings	listings	ShoeType	Shoe Type	VARCHAR(50)	N	N	Boots	Sneakers	N	N		
42	Listings	li	Listings	listings	Brand	Brand	VARCHAR(50)	N	N	asics	vans	N	N		
43	Listings	li	Listings	listings	Color	Color	VARCHAR(15)	N	N	black	yellow	N	N		

# Business Metadata

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Table	Data Domain	Criticality	Retention Policy	Security Classification	Data Steward									
2	users	Customers	Very High	Seven years	Internal	Customer Department Head									
3	creditcards	Customers	Very High	Seven years	Internal	Customer Department Head									
4	Orders	Orders	Very High	Seven Years	Internal	Order Department Head									
5	OrderShipments	Orders	Very High	Seven Years	Internal	Order Department Head									
6	Items	Inventory	Medium	Zero Years	Internal	Inventory Department Head									
7	listings	Listings	High	Two Years	Internal	Listings Department Head									
8	OrderItems	Orders	Very High	Seven Years	Internal	Order Department Head									
9	CustomerServiceRequests	Customers	Very High	Seven years	Internal	Customer Department Head									
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															
36															
37															
38															
39															
40															
41															
42															
43															
44															
45															
46															
47															
48															
49															
50															

	Technical Metadata	Business Metadata	Data Quality Issues	Standard Naming Conventions (St)	+
--	--------------------	-------------------	---------------------	----------------------------------	---

---

## **Step 3**

### Data Quality

#### Part 1: Profiling and Cleansing

# Data quality issues

	A	B	C	D	E	F	G	H	I	J
1	Existing Issue	Column	Table	Issue	Dimension	Description	Example	Suggested Resolution	Suggested Data Quality Rule	Data Quality Metric
2	Y	listingenddate	listings	Missing Data	Completeness	71.6% Null Values	listingenddate=11/1/20	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
3	Y	phone	customerservicerequests	Missing Data	Completeness	31.0% Null Values	phone=(237) 593-8419	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
4	Y	shoetype	listings	Missing Data	Completeness	16.0% Null Values	shoetype=Boots	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
5	Y	arrivaldate	items	Missing Data	Completeness	15.8% Null Values	arrivaldate=9/25/20	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
6	Y	email	customerservicerequests	Missing Data	Completeness	13.8% Null Values	email=bobby.vanderheyden@fakeemail.com	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
7	Y	orderid	customerservicerequests	Missing Data	Completeness	13.8% Null Values	orderid=914	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
8	Y	itemstatus	items	Missing Data	Completeness	13.6% Null Values	itemstatus=approved	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
9	Y	listingoldprice	orderitems	Missing Data	Completeness	11.5% Null Values	listingoldprice=10	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
10	Y	shippingaddress	orders	Missing Data	Completeness	7.1% Null Values	shippingaddress=0 Courage Drive	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
11	Y	color	listings	Missing Data, color is not consistent with the color in the items table	Completeness	2.8% Null Values	color=black	Implement Controls in the Application: add validation in the originating system to ensure that blank values are not accepted.	Check for missing values	% of missing values
12	Y	brand	listings	Missing Data	Completeness	2.3% Null Values	brand=ASICS	Refine Business Processes: establish a trusted source of truth (items table) and populate this information automatically for all records from that source.	Check for missing values	% of missing values
13	Y	sellerid	listings	The sellerid is not consistent with the sellerid in the items table	Consistency	99.5% of the sellerid are different given that the itemid is equal to productid	sellerid=121	Refine Business Processes: establish a trusted source of truth (items table) and populate this information automatically for all records from that source.	Check for conflicting values. Merge into one record with the correct value.	% of conflicting values
14	Y	condition	listings	The condition is not consistent with the condition in the items table	Consistency	26.0% of the condition are different given that the itemid is equal to productid	condition>New	Refine Business Processes: establish a trusted source of truth (items table) and populate this information automatically for all records from that source.	Check for conflicting values. Merge into one record with the correct value.	% of conflicting values
15	N	size	listings	The size can be inconsistent with the size in the items table	Consistency	0% of the size are different given that the itemid is equal to productid	size=0	Refine Business Processes: establish a trusted source of truth (items table) and populate this information automatically for all records from that source.	Check for conflicting values. Merge into one record with the correct value.	% of conflicting values

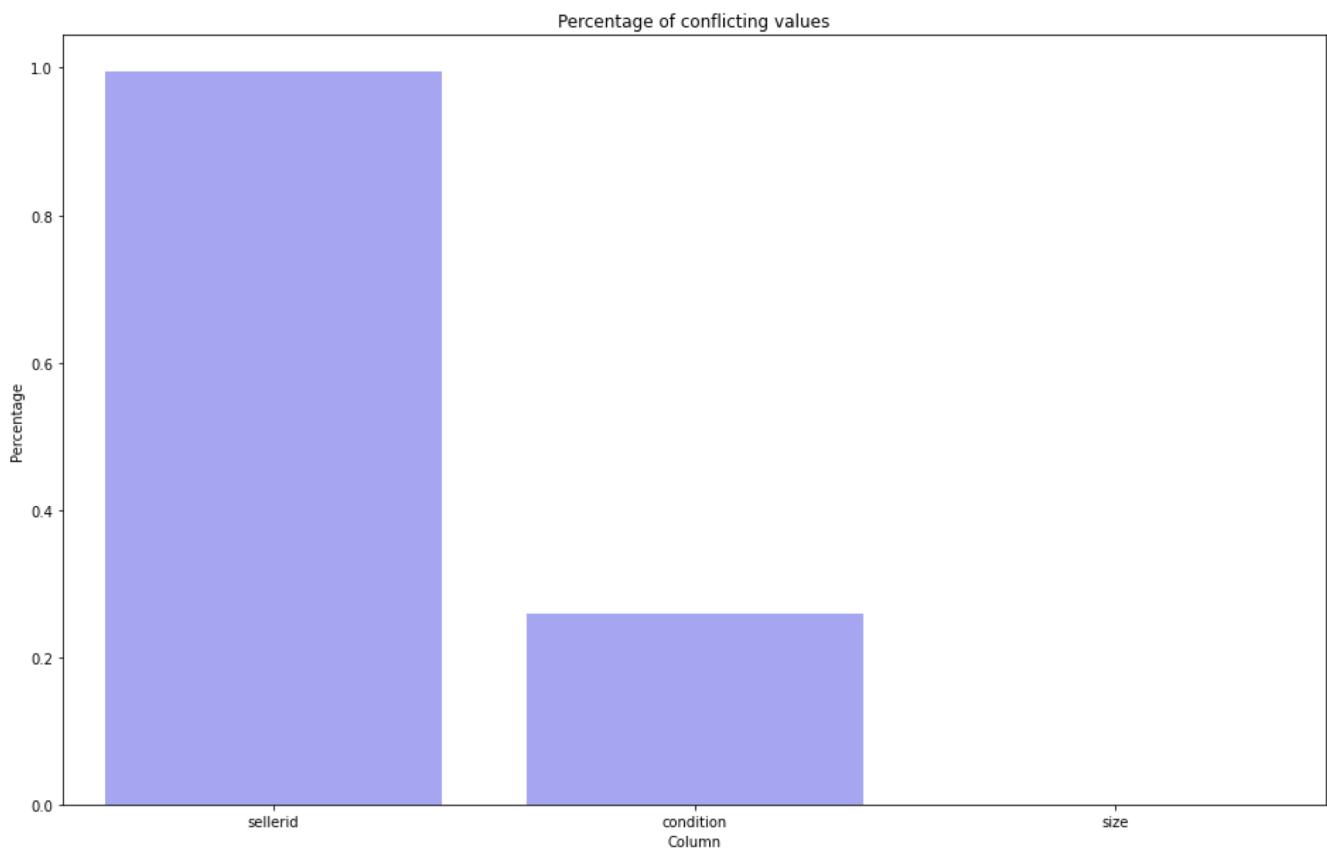
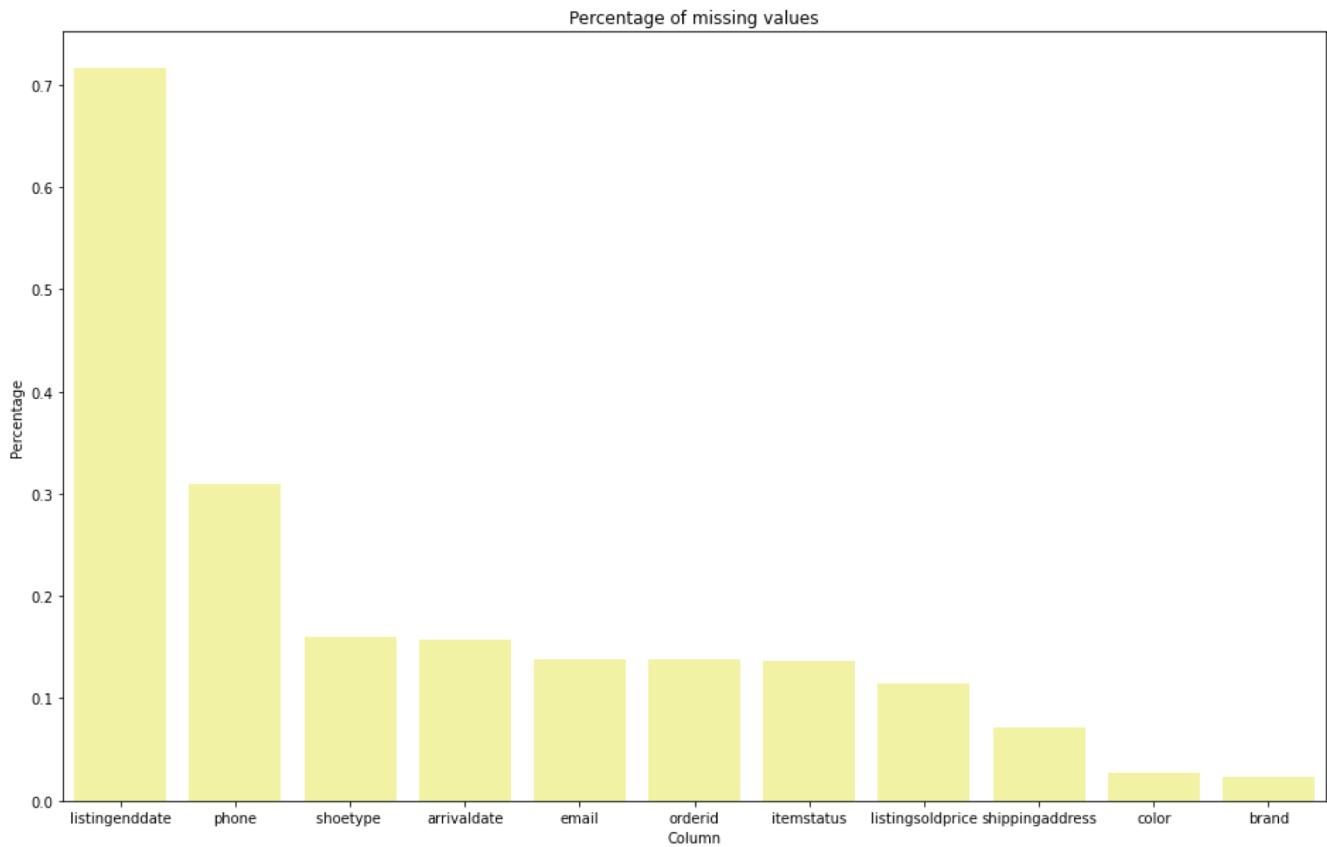
---

## Step 4

### Data Quality

### Part 2: Monitoring

# Data quality monitoring dashboard



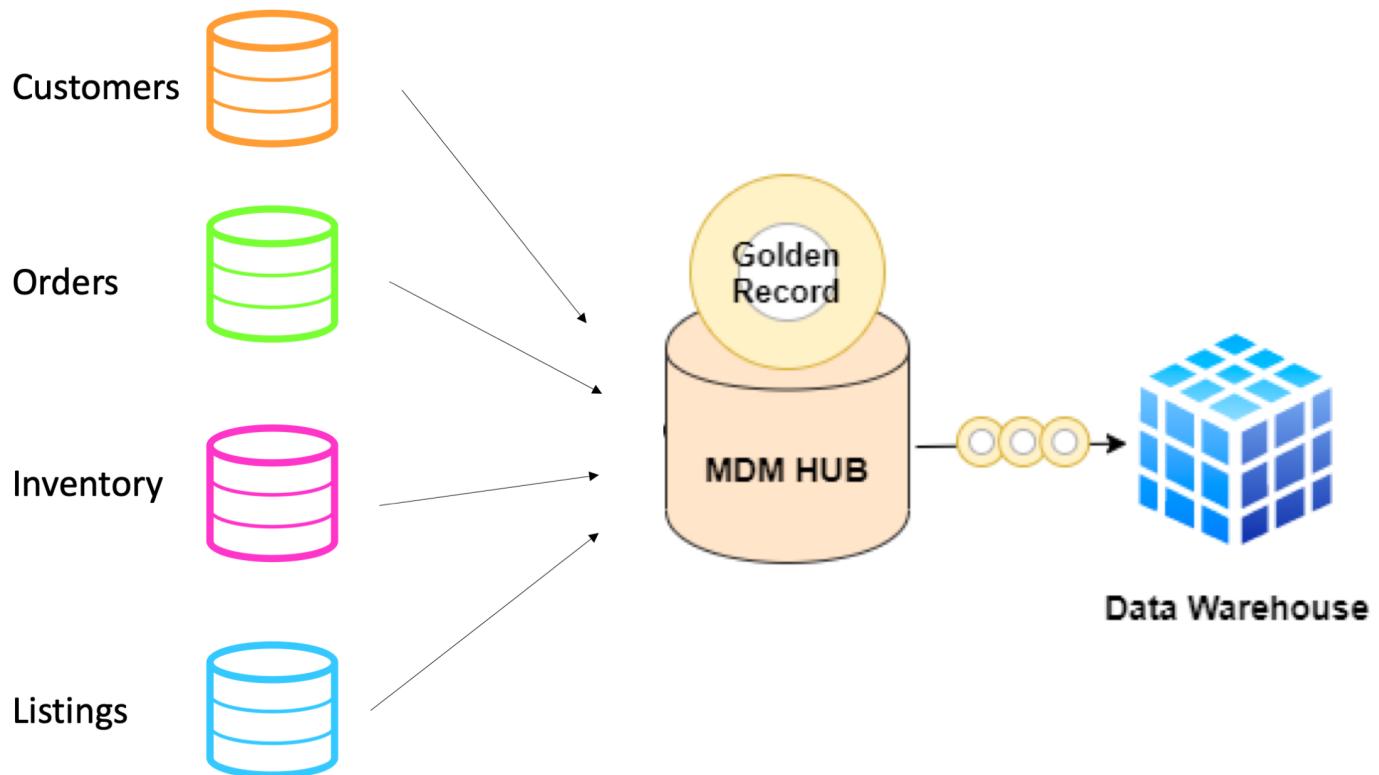
---

## **Step 5**

Master Data Management  
Part 1: MDM Architecture

## Proposed MDM implementation architecture

I've chosen to go with Consolidated style of architecture. SneakePark has a number of disparate systems that makes it difficult to trace the same item and seller across different systems. The scenario description states that the company plans to build out an Enterprise Data Warehouse solution to replace the existing legacy data warehouse, MS Access databases, and excel reports, which makes me think there are analytical needs. For that reason, I believe consolidating data from all these systems into a central MDM Hub that can feed merged, cleansed, and verified golden records to data warehouses and reports is the best solution here. In addition, this architecture is less invasive and has minimal disruption of existing systems.



---

## **Step 6**

### Master Data Management

#### Part 2: Master Record

**Matching rules** that will be used by the SneakerPark's MDM Hub to match item and customer entities between the company's different systems.

For Items:

- Match item data on itemid.
- Match item data on productid.

For Customers:

- Match customer data on userid.
- Match customer data on sellerid.
- Match customer data on creditcardid.

---

## Step 7

# Data Governance: Roles and Responsibilities

## **Data governance roles and responsibilities**

Data Governance Manager:

- Establish and govern an enterprise data governance implementation roadmap including strategic priorities for development of information-based capabilities
- Roll out an enterprise wide data governance framework, with a focus on improvement of data quality and the protection of sensitive data through modifications to organization behavior policies and standards, principles, governance metrics, processes, related tools and data architecture
- Define roles and responsibilities related to data governance and ensure clear accountability for stewardship of the company's principal information assets
- Serve as a liaison between Business and Functional areas and technology to ensure that data related business requirements for protecting sensitive data are clearly defined, communicated and well understood and considered as part of operational prioritization and planning
- Develop & maintain inventory of the enterprise information maps, including authoritative systems, owners

## **Data governance roles and responsibilities (cont'd)**

Data Quality Manager:

- Raising data quality and performance levels
- Providing a high level of data quality awareness across multiple staff profiles
- Proactively improving the quality of company reporting
- Evaluating and identifying where system enhancements are required

Master Data Manager:

- Design, construct, install, test and maintain highly scalable data pipelines with state-of-the-art monitoring and logging practices.
- Bring together large, complex and sparse data sets to meet functional and non-functional business requirements.
- Design and implement data tools for analytics and data scientist team members to help them in building, optimizing and tuning the product.
- Integrate new data management technologies and software engineering tools into existing structures.
- Help in building high-performance algorithms, prototypes, predictive models and proof of concepts.

## **Data governance roles and responsibilities (cont'd)**

Metadata Manager:

- Perform business requirements analysis for business rules and documents associated with population and extraction of metadata resources
- Develop, maintain, and support application, business glossary, business intelligence, data integration, data modeling, and database management resources
- Develop, maintain, and support Metadata Manager Security model
- Develop, maintain, and support linkage between Informatica Business Glossary and Metadata Manager assets
- Validate completeness of data lineage between the point of data element creation (system of record) and the point where it is consumed by business users (trusted data source)
- Monitor periodic loads of Metadata Manager resources to confirm successful load, indexing, and linking.
- Monitor linking completeness for Metadata Manager resources and troubleshoot issues with missing links.
- Provide ongoing support for metadata implementation, testing and integration
- Provide technical training and documentation in support of metadata policies and procedures

## Data governance roles and responsibilities (cont'd)

Data Steward:

- Define the data and identify assets within their own data domains. This ensures there isn't conflict with other data elements.
- Create processes and procedures along with access controls to monitor adherence. This includes establishing internal policies and standards—and enforcing those policies.
- Maintain quality of the data using customer feedback, concerns, questions; internally reporting metrics; evaluating and identifying issues; and coordinating and implementing corrections regularly.
- Optimize workflows and communications.
- Monitor data usage to assist teams, share best practice trends in data use, and provide insight into how and where teams can use data to help in day-to-day decision-making.
- Ensure compliance and security of the data. Data stewards are responsible for protecting the data—while providing information on potential risks and offering regulatory guidance.

Jake and Jessica have the skills required to fulfill the responsibilities of the data stewards. Sneaker Park will need to hire new employees to take up additional roles apart from the data steward.