1213bst is an online second-hand retailer who buys all of its product directly from consumers.
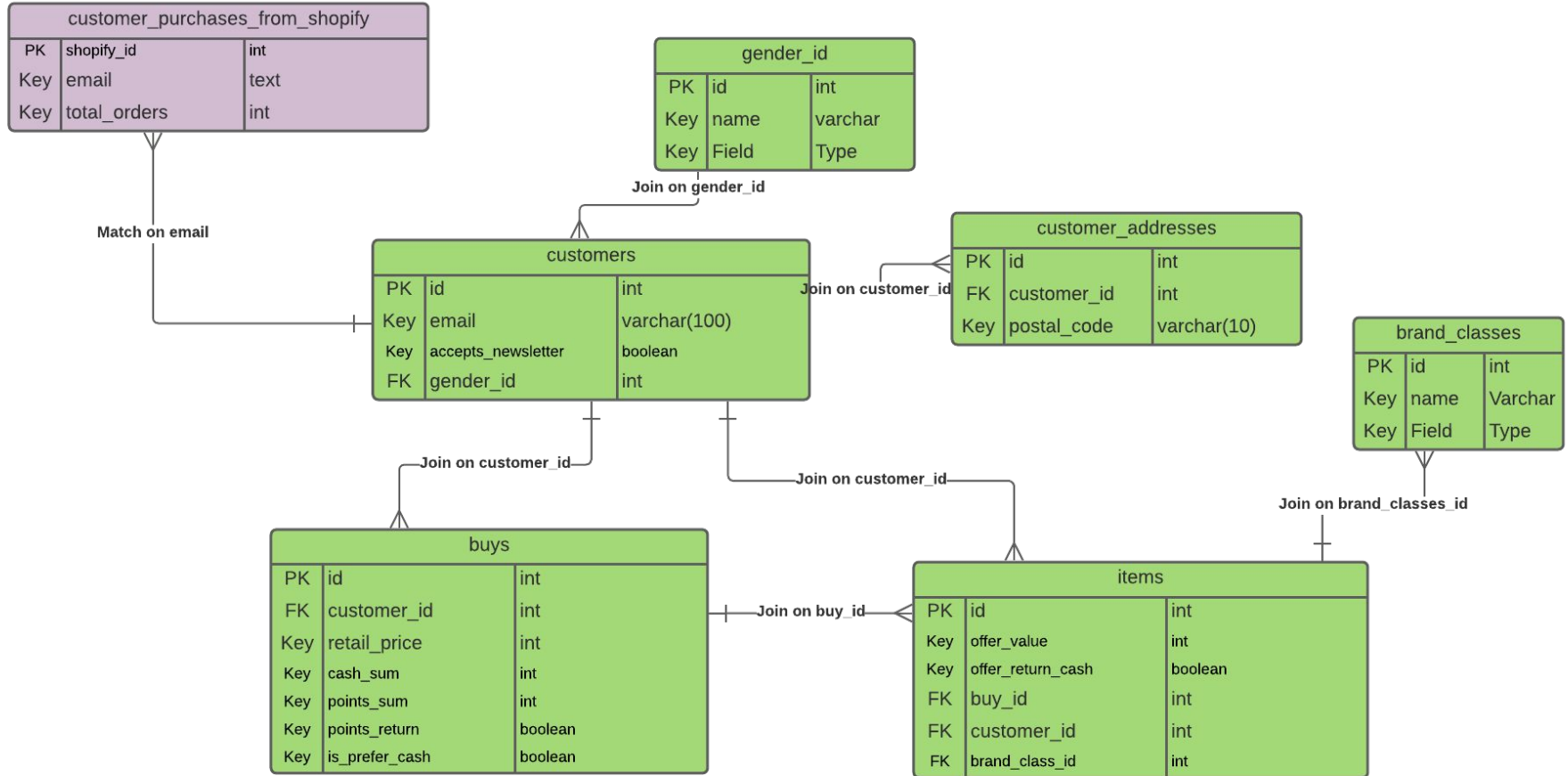
We offer consumers two payment options: **cash or trade credit**

Trade credit is essential both to our principles and our growth.

So, I wanted to build a model that can accurately predict:

Bryan,

do you prefer trade credit or cash?

trade   cash

Remember you get a lot more for trade credit. You can use it online an in-store for up to 2 years!

I first created an ERD of the main tables I wanted to use from our database

customer_purchases_from_shopify

PK	shopify_id	int
Key	email	text
Key	total_orders	int

gender_id

PK	id	int
Key	name	varchar
Key	Field	Type

Join on gender_id

customer_addresses

PK	id	int
FK	customer_id	int
Key	postal_code	varchar(10)

Join on customer_id

brand_classes

PK	id	int
Key	name	Varchar
Key	Field	Type

Join on brand_classes_id

customers

PK	id	int
Key	email	varchar(100)
Key	accepts_newsletter	boolean
FK	gender_id	int

Match on email

Join on customer_id

Join on customer_id

buys

PK	id	int
FK	customer_id	int
Key	retail_price	int
Key	cash_sum	int
Key	points_sum	int
Key	points_return	boolean
Key	is_prefer_cash	boolean

Join on buy_id

items

PK	id	int
Key	offer_value	int
Key	offer_return_cash	boolean
FK	buy_id	int
FK	customer_id	int
FK	brand_class_id	int

I cleaned and wrangled, bucketed, binned, and encoded and in the end ran 5 different models.

Model	Accuracy	r2
Logistic Regression	0.805	0.403
KNN	0.785	0.120
Decision Tree	0.259	0.259
Random Forest	0.389	0.389
Decision Boosting	0.459	0.459

- Which model is telling me the truth?
- Why are the results to wildly divergent?
- How could future iterations make our outcome better?
- How can take this an use it to inform business decisions?