1 2
1 3

b s t

BUY
SELL
TRADE

#howfashionshouldbe

Will you choose Cash or Trade?

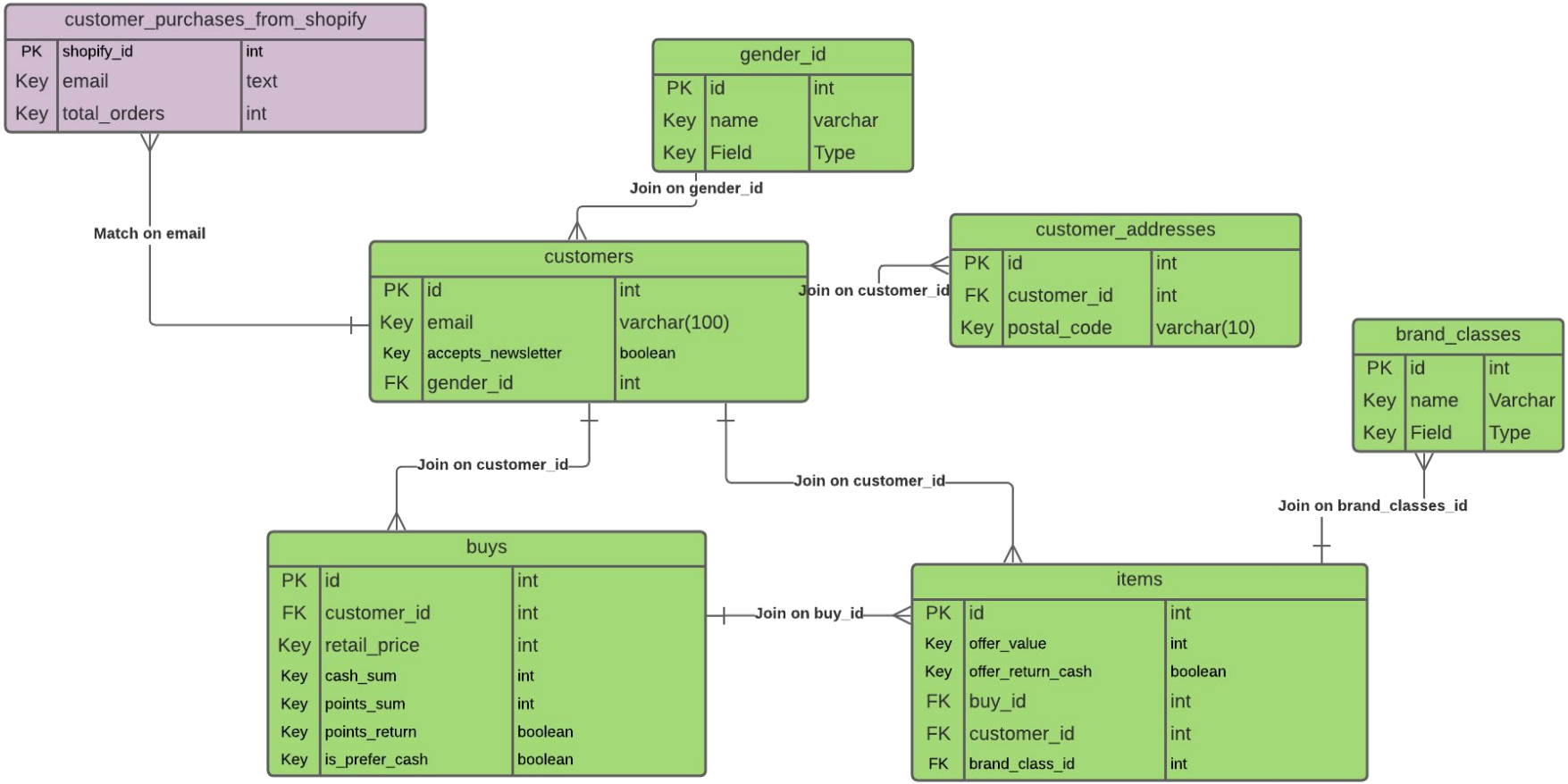
1213bst is an online second-hand retailer who buys all of its product directly from consumers.

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:

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



My initial dataframe

	buy_id	customer_id	chose_trade	total_offer	prefer_cash	total_items	postal_code	gender	brand_class_list	total_buys	buying_customer	trading_customer
0	1	660	False	321.0	NaN	5	10178	None	Missing brand class,Missing brand class,Missin...	2	False	False
1	2	661	True	42.0	NaN	3	NaN	None	Missing brand class,Missing brand class,Missin...	1	False	False
2	3	662	True	90.0	NaN	3	UNKOWN	None	Missing brand class,Missing brand class,Missin...	1	False	False

My final dataframe before running my model

	chose_trade	buying_customer	trading_customer	total_offer_new	total_items_new	total_buys_new	postal_code_10115	postal_code_10119	postal_code_1017
0	0	0	0	5	3	1	0	0	
1	1	0	0	1	1	2	0	0	
2	0	0	0	5	4	4	0	0	
3	1	0	0	1	1	2	0	0	

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?