Data Cleaning



	Dress_ID	Style	Price	Rating	Size	Season	NeckLine	SleeveLength	waiseline	Material	FabricType	Decoration	Type	
0	1006032852	Sexy	Low	4.6	М	Summer	o-neck	sleevless	empire	NaN	chiffon	ruffles	animal	1
1	1212192089	Casual	Low	0.0	L	Summer	o-neck	Petal	natural	microfiber	NaN	ruffles	animal	0
2	1190380701	vintage	High	0.0	L	Automn	o-neck	full	natural	polyster	NaN	NaN	print	0
3	966005983	Brief	Average	4.6	L	Spring	o-neck	full	natural	silk	chiffon	embroidary	print	1
	076220E41	outo	Low	4.5		Cummor	o nook	buttorfly	notural	obiffonfobrio	ohiffon	how	dot	0

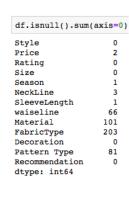
Data exploration



Correct typos "waiseline" to "WaistLine", etc.

Uniform characters — Lowercase categorical columns

NaN values In 500 rows



WaistLine, Test model scores
Material, →66-203 missing → with or without
these columns
PatternType

Price,
Season,
1-3 missing
NeckLine,
SleeveLength

Fill with most common values

Treating each column

Rating — Ratings can't be zero — Remove

Price "Average" to "Medium"

("cap-sleeves", "capsleeves")

("half", "halfsleeve")
("sleveless", "sleeveless")
("sleevless", "sleeveless")
("sleevless", "sleeveless")
("sleevless", "sleeveless")
("sleeveless", "sleeveless")
("thressgatar", "threequarter")
("urndowncollor", "turndowncollar")
("turndowncollor", "turndowncollar")
("leapord", "leopard")

Reduce Levels

Style

("party", "special_occasion")
("novelty", "other")
("bohemian", "boho_vintage")
("vintage", "boho_vintage")
("fashion", "other")
("brief", "casual")
("cute", "casual")

("bowneck", "bow_and_ruffle")
("ruffled", "bow_and_ruffle")
("slash-neck", "slash_and_boat")
("boat-neck", "slash_and_boat")
("sweetheart", "sweetheart_and_square")
("sqare-collor", "sweetheart_and_square")
("backless", "other")
("peterpan-collor", "other")
("halter", "other")
("open", "other")
("scoop", "other")

SleeveLength _____("turndowncollar", "full") ("butterfly", "halfsleeve") ("capsleeves", "short")

("chiffonfabric", "chiffon")
("shiffon", "chiffon")
("milksilk", "silk")
("model", "modal")
("wool], "wool_and_cashmere")
("cashmere", "wool_and_cashmere")
("knitting", "knit")
("spandex", "knit")
("lycra", "knit")
("modal", "cotton")

.