

orders_data

PURCHASE_TS

- ▶ Changed all dates to MM/DD/YYYY for format consistency
=DATE(YEAR(C2), MONTH(C2), DAY(C2))
- ▶ Added columns for PURCHASE_MONTH and PURCHASE_YEAR
=DATE(YEAR(C2), MONTH(C2), 1) =YEAR(C2)

REFUND_TS

- ▶ Removed nonsensical dates (dates in the future)
=IF(OR(K2 = "", K2 > TODAY()), "", K2)
- ▶ Added binary column for REFUNDED
=IF(K2 = "", 0, 1)

PRODUCT_NAME

- ▶ Corrected product names for format consistency
- | <u>Inconsistent Format</u> | <u>Reformatted</u> |
|----------------------------|----------------------------|
| 27in 4K gaming monitor | 27in 4K Gaming Monitor |
| 27in"" 4k gaming monitor | 27in 4K Gaming Monitor |
| bose soundsport headphones | Bose Soundsport Headphones |

MISSING AND NONSENSICAL DATA

- ▶ PURCHASED_TS had 3 missing values. Remained since < 1% of data
- ▶ DAYS_TO_SHIP had 27 nonsensical values. Cleared since < 1% of data
- ▶ DAYS_TO_RETURN had an average refund time of 740 days.
- ▶ USD_PRICE had 158 '0' values. Remained since < 1% of data
- ▶ COUNTRY_CODE had 206 missing or nonsensical values. Remained since < 1% of data
- ▶ COUNTRY_NAME had 206 missing values. Remained since < 1% of data

AUGMENT DATA

- ▶ Added columns for DAYS_TO_SHIP & DAYS_TO_DELIVER
- ▶ Added column for DAYS_TO_RETURN
- ▶ Imported REGION column via XLOOKUP
- ▶ Imported COUNTRY_NAME column via XLOOKUP

country_lookup

REGION

- ▶ Imputed values for missing region codes

<u>Country</u>	<u>Sources</u>
BJ	ISO CIA World Factbook
BM	ISO CIA World Factbook

- ▶ Replaced 'North America' and 'x' with 'NA' for format consistency
- ▶ Imputed Country Name column via XLOOKUP

MISSING AND NONSENSICAL DATA

► COUNTRY_CODE had 10 nonsensical values. Remained since < 1% of data

