Data Cleaning

skuinfo: packsize

To avoid type error when copying the skuinfo.csv to database, the field `packsize` was set as text type. By running the SQL query below, there is some rows with `packsize` that cannot be converted by integers automatically.

```
SELECT *
FROM skuinfo
WHERE packsize !~ E'^\\d+$';
```

So we are going to replace those values by the mode of `packsize` : 1.

```
SELECT mode() WITHIN GROUP (ORDER BY packsize::integer)
FROM skuinfo
WHERE packsize ~ E'^\\d+$';

Output:
1

UPDATE skuinfo
SET packsize = '1'
WHERE sku IN (
    SELECT sku
    FROM skuinfo
    WHERE packsize !~ E'^\\d+$'
);
```

Then we can convert text to integer.

```
ALTER TABLE skuinfo
ALTER COLUMN packsize TYPE integer USING packsize::integer;
```

trnsact: salesdate

Change the data type of `salesdate` from text to date.

```
ALTER TABLE trnsact
ALTER COLUMN saledate
TYPE DATE USING saledate::date;
```

Build index on `salesdate` for faster CRUD operations.

```
CREATE INDEX saledate_idx ON trnsact(saledate);
```

trnsact: quantity

Build index on `quantity` for faster CRUD operations.

```
CREATE INDEX quantity_idx ON trnsact(quantity);
```

Change the data type of quantity from text to integer.

```
ALTER TABLE trnsact
ALTER COLUMN quantity TYPE integer USING quantity::integer;
```