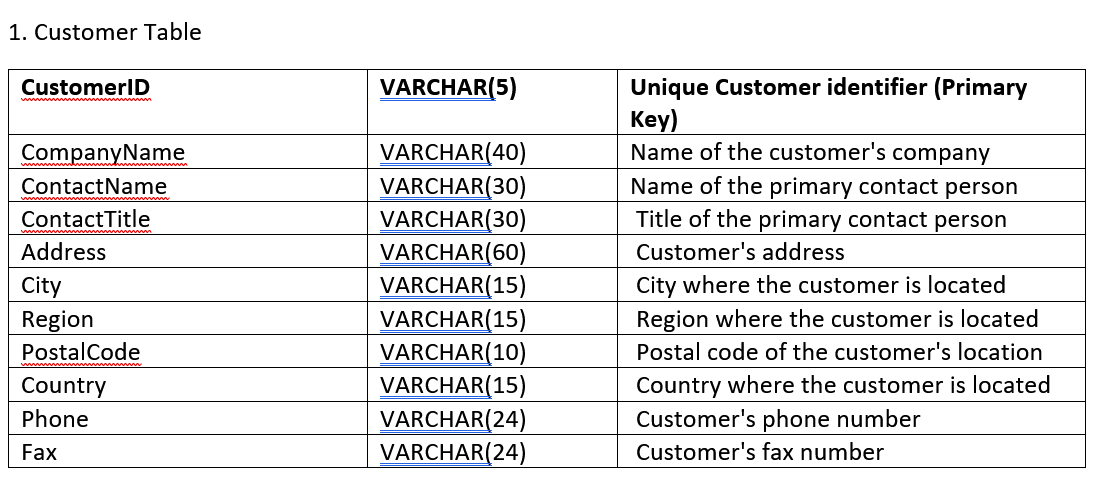
Customer Table

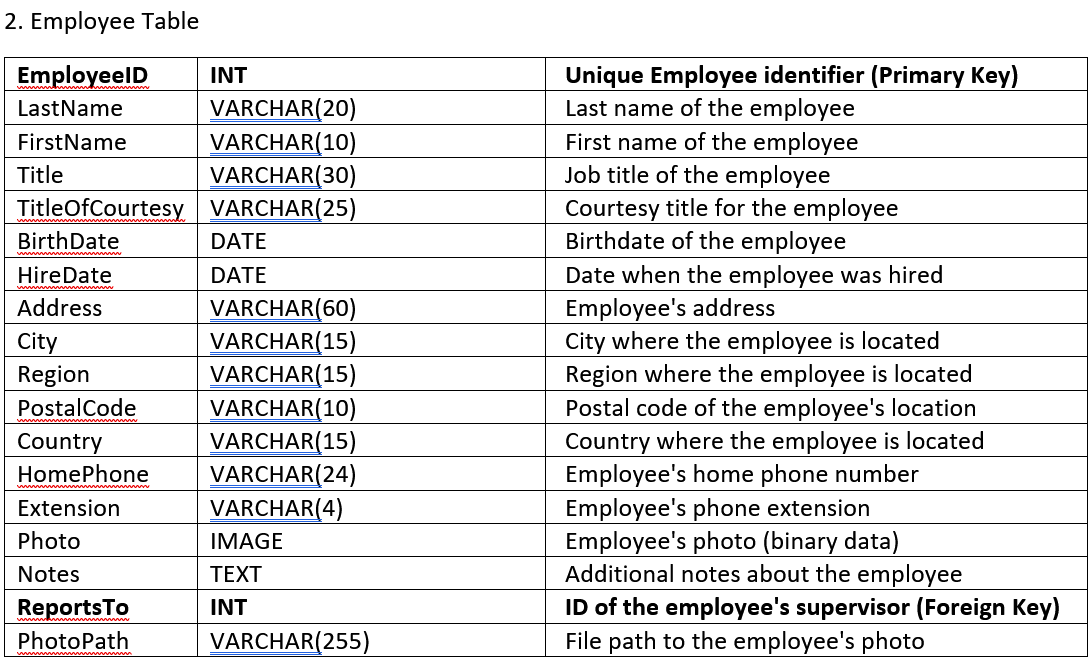


* 91 entries
* NaN in region, postalCode, fax
  + NaN postalCode at customerID HUNGO
* Verifying VARCHAR information stated in Data Dict:
  + Highest VARCHAR of City exceeds 15, highest being 17. (To be cleansed, incorrect input of City being ’12Rio de Janeiro’ while ‘Rio de Janeiro’ already exist in the dataset.)
  + [CHANGED]
* Customer phone number can be the same as fax number.

NO duplicates in CustomerID (Unique Identifier)

* Duplicates in contactTitle, city, region, postalCode, country, fax. (Valid)
* postalCode can be the same but with different address input.

Employee Table



* 9 entries
* NaN in region and reportsTo
  + employeeID 2 is a Vice President of Sales, thus ‘reportsTo’ is NaN.
* No duplicates in employeeID (Unique Identifier)
* There are duplicates in both ‘photo’ and ‘photoPath’, indicating that perhaps the employees are using some default profile pictures??
* Datetime format shown is “YYYY-MM-DD HH:MM:SS.SSS”

Order Table

A list of orders

Description automatically generated with medium confidence

* Python recognized that there are 830 entries of data but only max of 800 rows of data are present. To validate that, there are also empty rows of data for order dataset seen in GitHub.
* Ignoring these additional 30 entries of NaN rows, NaN are present in ‘shippedDate’, ‘shipRegion’, ‘shipPostalCode’. Majority of the NaN resides in ‘shipRegion’ (490/800 entries).
* Datetime format is in “DD-MMM-YYYY’
* No duplicates in orderID (Primary Key and Unique Identifier of this table)
* No issue with the foreign key of customerID and employeeID, count of unique is valid according to other tables.

Order Detail Table

A close-up of a person's face

Description automatically generated

* 2062 entries
* No NaN entries for this table.
* Number of unique orderID corresponds to the number of unique orderID in ‘order’ table excluding the NaN entries.
* Discounts are in decimals. 0.00, 0.05 (5%), 0.10(10%), etc. Highest discount being 0.25 AKA 25%.
* Number of unique ‘productID’ corresponds to the one in product table.

Product Table

A list of product names

Description automatically generated with medium confidence

* 77 entries.
* No NaN for this table.
* No duplicated ‘productID’, all entries of unique products.
* Number of unique ‘supplierID’ corresponds with the count in supplier table.

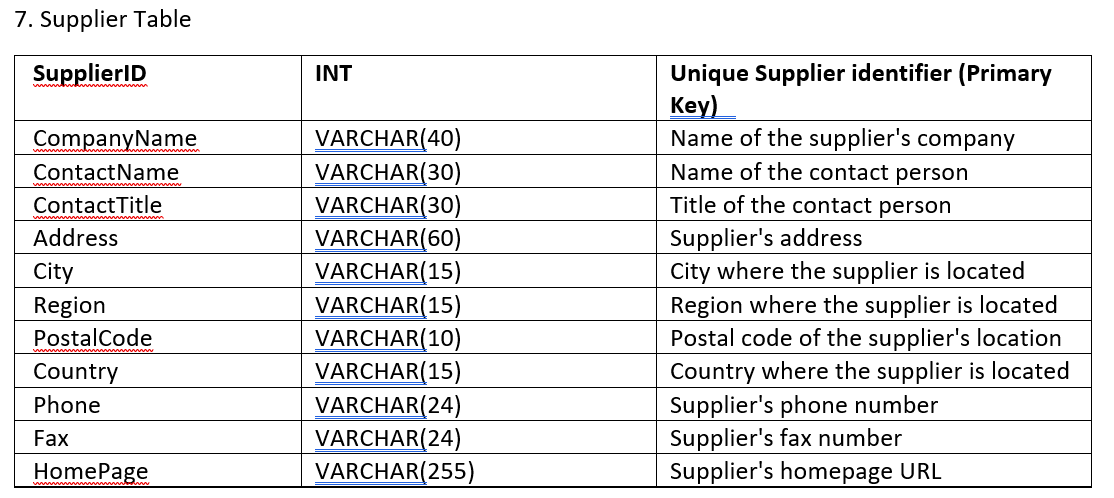
Shipper Table

A black line on a white background

Description automatically generated

* 3 entries.
* No NaN.
* Shippers from country called El Salvador in Central America. Country code 503.

Supplier Table



* 29 entries
* NaN in ‘region’, ‘fax’, and ‘homePage’.
* Duplication of fax number and homePage url, is that valid??
* No duplication of ‘supplierID’ (Unique Identifier)

Category Table

A screenshot of a computer

Description automatically generated

* 8 entries.
* No NaN.
* Duplication of picture of category used in data.
* They all use the same image.

EmployeeTerritory Table

A close-up of a sign

Description automatically generated

* 49 entries.
* No NaN
* Number of unique ‘employeeID’ corresponds to the count in employee table.
* All territoryID are unique.

Region Table

A white background with a black line

Description automatically generated with medium confidence

* 4 entries.
* No NaN
* No issue here!

Territory Table

A black text box with black text

Description automatically generated

* 53 entries.
* No NaN
* No duplicates of ‘territoryID’ but there are duplicates of ‘territoryDescription’.
* A screenshot of a computer

  Description automatically generated