Data Quality and Performance in Action

1. Data Quality Assessment and Improvement

Dataset Selection and Initial Assessment

For this project, I selected an Excel sample dataset which contains responses from a survey conducted in 2024 to understand gaming habits, preferences, and player behaviours. The dataset provides valuable insights into the gaming landscape, helping developers, marketers, and researchers identify trends and preferences in the gaming community. The data was assessed based on five quality dimensions: accuracy, completeness, consistency, timeliness, and uniqueness.

Assessment

- **Accuracy**: Data was cross-verified with known valid ranges.
- **Completeness**: Some columns have up to 40% missing values, especially those with duplicate headers (columns from N to V are duplications of columns E to M), as seen in Figure 1.
- **Consistency**: Formatting inconsistencies were noted, especially in the 'Location' and 'What is your favourite game?' fields.
- **Timeliness**: Timestamps are within a reasonable range (October 23–25, 2024), indicating data was collected over three days.
- Uniqueness: No duplicate rows were found in the dataset.

Data Quality Improvement Steps

I focused on improving completeness and consistency. The steps were as follows:

- 1. **Completeness**: Removing duplicated columns (from N to V).
- 2. **Consistency**: Selecting one format for 'Bangalore, Karnataka' (as seen in Figure 2) and 'Bhubaneswar, Odisha' (as seen in Figure 3) on 'Location', using Capital first and State second. And one format for 'Call of Duty' on 'What is your favorite game?' (as seen in figure 4).

Metrics: After cleaning, consistency was improved to 95%, and duplicate entries were entirely removed (see Figure 5).

Screenshots and Annotations

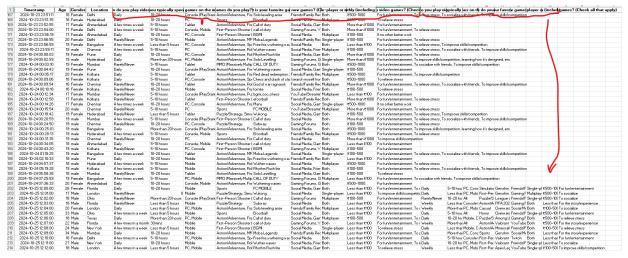
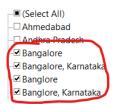
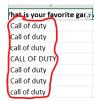


Figure 1 shows the initial state of the data, with highlighted areas indicating missing values and duplicates.







Figures 2 and 3 show the inconsistent data, on 'Location' field.

Figure 4 shows the inconsistent data on 'What is your favorite game?' field.

| | | Which device draws play sense as the mart! (Check of that - | | | | De yeaprafar ringkraleyor armobial ayun ad v | Haumoch da yauraund en arming mentilly (including in-game purchasor, nou army | Mile da yang lier video generi (Check off that apply) |
|--|----------------------------|--|---|--------------------------|---|--|---|--|
| ertathings 20 Male Empelers, Comptable Schoptin | rinmarch Larthurftmar | Makin | | | Sanial My dia | Singlewinger | Larr than F 900 | Tending charge |
| erracers to their Marshall Daily | 11-21hany | Makita | ArtistAdvertore, 1910 (Hastinal) Hubipleson Orline) | Makakenende | Ericolet's with European Artists | Mehidasar | Last than FMR | Facilitation technological and Caracia Sacrada Asianda, Taingram of Hot competition |
| erracycle W Mely Shabeaurum, Odoby Advoysia | ringuek 5-95ppr | POPMAN | Action/Advanture, Sports, HHO (Massingly Multiples or On- | Free fire outbering over | Societ Me So | Buth | | Factor Festivation and |
| errorge & Male Balancer Odeka Dally | Th-Difference | Media | Artist/Advanture, Sale-Playing Banas (FPS), PR92 (Plan | | | | | Facility Europe to Santage and Talmace and Eleforation things |
| erraticite Il famale Empelere, Comptabe Schaplin | | PERMAN | | | | | | Tendensions. |
| erraticide il famelo Andirefredata Dally | Learthenfilmer | Canada (Phas Station, Share, etc.), Philidia | Action/Admintore, First Petron Shorter (FPS), PanalotSt. | | | | | |
| | | | | | | | | Fairfue Funtum element, Tarro Europetruer, Tarro inflas with friunds, Taingreen aski |
| 4FFRQ124 17 Famala Bengalara, Kemataka Daily | Loss than 5 keyes | Course (PlayStation, West, etc.), Handfold devices (Naturale | | | | | | Tendenormer |
| errordide 20 Male Bhabesonum, Odinka Afronia | | PC/Nets | Action/Advanture, First Paren Stantor (FFS), Parel of St. | | Social Madia, Gaming Forums, Friendoff anally Recom- | | | Ferhalphynaineun |
| erraticité de Maio Bungalere, Camataka Afrontin | rinappet 5-90 kmps | | Articulturature, Einst-Parren Stanter (TPS) | Ryd deader-from time? | Friends/Family Recommendations | Path | Last than 7 900 | Farrhantentuntainment, Taroformetruer |
| | | | | | | | | |
| erracycoa M Male Bengelers, Kemoteko Delly | Mary than 26 heavy | PG, Gauralo (PlayStotian, Shar, etc.), Plabilie | | Gedelum requestion | Friendry enity forcement detical, for fish of Street | | | Farrha Fentuntainment, Tarre Servertruor |
| erracios W. Male Empelers Comptable Streetin | | | | Faraba | Secret Marker Friands Warnity Resembles Artists | | | Fartherfuntationant, Territoryateur, Telepareur tiferteenauthien |
| | | | | | | | | |
| errorizate di Fomelo Bungalero, Komentako Arlosarian | | PO,HARL | ArtistAdvartors, Possis / Stratory, Shookeline (e.g., The | | | | T100-510 | Facturization and Tarafavorium, Taracidiae with friends, Taingenner |
| errandiche ff Malu Bungelere, Kamoneka Dally | 5-90 hover | PC,Pbblk | | | | | Location F100 | Factor Festivation and, Taro Security ar |
| SPERIORDA 25 Hale Bengelers, Kamoteko Daily | 9-Wileser | PEPARA | Action/Adventure, First Forest Stantor (FFS), Aniv Physics | | Social Hudia, Ganing Forums, Game Revision (Blogs, | | Lorrithan F 900 | Ferhalfunkunteismunk, Teimpressunklärfrempunktion, Leaninghous Hir duries. |
| | | | | | | | | |
| OFFECCIO 99 Famala Sympolery Especials Symplest | | 1 Maile | | Walkers | | | | Tainannuch Bhirannuchlin |
| Printings W. Functio McGrand Computers Sandoff | | Mate | | OutletDate | | | | Factor fundament |
| | | | | | | | | |
| SPERISES II male - Bengelers, Kemotoko Keroly Pf | | Carvelo (FlayStation, Sharpate.) | Persidibuter | Selver | | | | Familia Festuritain ment |
| 2018/2024 10 Famala Bengalara, Kamadaka Daily | Lase than 5 house | Makika | Artisal Managery, First Parent Stanton (FPS), Rate Physi- | | | | | Farrication telement, Taroforcerous, Taranislas with triang, Taisagrans |
| 2498Q4Q4 99 Famelo Bungelers, Kamataka Adagrim | ringget LarthmStour | PO.HARA | Senv | C/earth-off | Societty-Sa | Dath | Lorenthan 7 900 | Facility Fernandian por |
| principle to functio Bengatery Compatable Dally | Many than 25 heavy | Meta | Action Address to a Contraction Shows at 177 St. Rate Photo | (Call at Date) | Sprint Hydra, Spring Forums, Some Surface (Street, | | | Partie fest of classest, Tarofacourtsur |
| PRESIDE TO Family Join University Afrontis | | PCPMA | First Farran Shanter (FFS), Rain Playing Some (RFS), St | | | | | Farther fundantainment, Terreferouritrum, Terrefelles with Ariando, Teingrens |
| | | | | | | | | Tanahanatan |
| 2018/2024 20 male Brongsberr, Komotoko Africa tim | | Nells | | | | | | |
| OFFRESCO 91 malu Bungalara, Kamataka Dally | 19-24 Augu | MAN | Action/Advanture, 1990 (Hestinal) Hubbilanus Online) | Hobelsonde | Friends/Fenilly Recommendations | Highleur | Location 7900 | Fairfus Fentumeium ent, Taxociellas with friends, Taimprone skillef competitio |
| promoting to Family Marghall Shoulder | ringual 9-90 kmm | Mate. | Action/Advanture, Eponte, HHID (Massicule Phillips over the | Free Procumbaringumour | Secial Marks | fast. | Larrithan F 900 | Farchar Fanhartainmanh |
| | ti-litary | | ArtistAdvanture, Sale-Playing Samer (SPS), 1990 (Haw | | | Bath . | Lear than 7 100 | Factor fundamental Taimprens deliber competition |
| OTROPO IT Male Stell-marrow Origina Advanta | vin succh Laur than Shaper | Melic | Action/Advanture, Falls Planing Sense (RPS) | Rhythm Ruck Sty | | Dath | Lear them 7 100 | Tenfraguer |
| 1990004 C functs Smarker Smarks Dally | Larries Steam | PO PIATA | | | | | | Farthe Funtumalement, Terra Europement, Terra Gallerian (A. Telenoment |
| | | | Action/Adventure, First Petron Shorter (TPS), Paggle/St | | | | | |
| SPEANING IT Male dediceFredat Daily | Lord No. 9 hours | Carvele (PlayStation, Barr, etc.), Plobile | PRINCH aminuty Haltiplepur Online) | Celebry | | | | Tendinos/mer |
| 1988/2024 20 Hale Bungelers, Kamadaka Africa bin | | | derinablehandere, Finds Person Stanter (FPS), Panaloffic | | Social Hardin, Gaming Forema, Friendoff anally Recom- | | | Farrhadenhartainment |
| CHROSE William Statement Office Abouting | cinquest 5-99 tags | PC.PMIN | Action/Advanture, From Pages Stantar (FPS) | | Friends/Femily Recommendations | Dark | Lagration (199) | Farthe Lenturtainment, Tarreforeurtruck |
| Principal M Famala Banadera Kamataka Situatin | | PO | | | | | | Prestharhamarustk |
| printings William Emarker, Comptabe Dally | Mary than Jifferson | Media | | Sadelumragnere) | Friends/Family Recommendations, Smith but Street | | | Fartha (enhantainment, Tarafavorituer |
| | | | | | | | | |
| 2018/2024 W Famela Bungslere, Kamataka Afaratin | | PC, Conzelo (PlayStetien, Illien, etc.), Hebite | | Fernite | | | | Fartus fentuntainment, Taro Escostruor, Taimpromo et ille frança et idas. |
| 2018/2024 19 Malu Bengelers, Kemeteka Afroptin | | PO | Action/Advanture, Puzzle/Stratury, Sincletian (e.g., The | | | | | Factor Festivation and, Tanofesser truez, Taracinities with friends, Taingeans. |
| SPENDEDE 19 Famelo Bengelers, Kamateka Dally | 9-90 hover | PEPMA | Flort-Furran Shastur (FPS), Spertr | ofestioli | Friends/Femily/Furcemeus-Sotiesz | East | Lorriton/100 | Ferning Fundant element, Terrolloss et mer |
| | | | | | | | | |
| STRONG W Famala Channel Daily | 5-Whent | Cavado, Habile | | | | Singlephoer | | Factor fest establishment, Tarrefessertruer |
| MRDIDE W. Fundo Estado Dalo | Mary they Different | Caucalo Habile | | Waters | | | | Tainannuckilly/canauthing |
| | | | | | | | | |
| SYSMOTOR SI Famelo Pano Delly | 5-10 keyer | Cewelr, Hebits | Flort-Furran Shaptur (FPS) | CalletDay | | | | Ferrise Fundamination and |
| 241842424 11 Famels Brooplers, Konstaba Daily | Lose than 5 house | Carrelo (FlayStation, Shan, etc.) | Pasalofficulary | Saury | | | | Facilita Estimataina est |
| | | | Action/Advanture, First Petron Shorter (FPS), Rafe Physic | | | | | Factor fentionalisment, Tarreferouttrust, Tarresinias sektotriands, Taingrans |
| | Lore than Shower | | | | | Heliphour | | |
| Printege W Famels Empelor Comptain Stuplin | | Canada, Matilia | Series/Advanture, First-Parrey Disease (FPE), Rate-Plant | | Special Markins Special Foreign Special Physics (Edward Stew) | | | Farthe Funturinian and Terralian votices |
| | | | | | | | | |
| 3HM2624 26 Male Pone Africania | | P0,0warls | First Parson Sheater (FPS), Robe Playing Somes (RPG), Si | | Gaming Farence, Partick of Strumping glotterne (Tuit | | | Fartus funtum sina unt, Taro Euro struur, Taraninilau with friunds, Taingman |
| OFFRIDIDE 91 Male Channel Delty | Mary then 20 has u | Hakkir | Flort Pysian Shaptur (FPS) | | | | | Tenfewerner |
| SHROUND K. Halo Kalketo Kerulyfft | our Larthon Shower | PG Coursely | Action/Adventure, PRIO (Matricula Philiplean Caline) | Hobelsquade | | | Hara-than F8968 | Ferhantsotunteinmont, Terecielles with Arisode, Teimpress chillet computition |
| SYSSESS Si Famela Pana difusion | r march Mary than 20 harve | PC | Series/Advantury, Specie, MMO (Plansingly Phillips over Dr. | | SerielMedia | Eath . | | Ferhalistationant |
| OrtalDigit 17 Male Medicated Adoption | | PO, Coursely | Action/Advanture, Sale-Playing Source (SPS), 1990 (Plan | | | | | Farthafestertainment, Taimerenschillefeangerbien |
| SHADON IT Help Page Daily | Trittany | PC.PhA.b. | | | | | | Tendengener |
| | | Teles | | | | | | Tambignormum Factor (anticipation and Tambignormum Tampidity with triangle Taignorm |
| | II-24heav | | Action Advanture, First Person Shouter (FPS), Parallelle | | | | | |
| 2018/2024 20 Famela Kalketa Randy/D | | Carvels (FlayStation, Shan, etc.) | | Callabbay | | | | Tendingstow |
| SHRQIQE 91 Famels Page Daily | Many through Sancy | Telifort | Action/Advanture, ComPages Shorter (TPS), Page 1/15) | Workeringsprove | Social Hydra, Ganing Forumy, Friendoff and a flucture | Duck | Many than F1009 | Factor Leanura in month |
| principes in Famala Bengalera Kemataka Dalla | Many than 26 heavy | PG.Coverly | | | | Highland | Three more | Facility Control (San Annual Control |
| MARGICA William Marchel Daily | 9-Wilanasi | Teklet | | | | Basicalmer | | Propoblerhettersonk |
| 2198/2124 9 Famels Stimulated Resulpti | | Metit | | | Friends Camille Recommendations, Smills of Street | | | Factor Letter Land Terreformations |
| remarks or ranked minurished Randy/N | | | | | | | | |
| MINDER Miller Meterolat Resolves | | Carvelo (PlayStation, liber, etc.) | Action/Adventure, Flort Presen Sheater (TPS) | Fernits | | | | Fairfue Fronzistalem von, Tano Errostrucz, Talimprovo z killoficampot Nies. |
| IMBRORDE M Famels Manhal Resulph | | Tellist | ArtistAdvanture, PumfetStrategy, Shookeline (e.g., The | hami, one, cheer | | | | Farrhan Funductions and, Tarrella vocatiouss, Tarrectellas with Ariando, Talangean |
| PRODE W. Famala Empelore, Kamataka Randy R. | nor Shithaw | PC, Courselo | First Fusion Sheeter (FPS), Specie | efective. | Friends/Family/Farences-detiens | | | Fertherfenhantelement, Terreferoutiner |
| STRICTOR 19 Male Pone Advanta | | PO | Action Midward are, Front Papers, Shouter (FPS), Raly Phon | Mana | Social My dia, Gamina Forume, Game Review (Blow. | Buth | Lagration (1900) | Tenfrontur |
| PRINCIPAL OF Famels Channel Resolver | | Table | | | | | | Tanolis vertrum, Far fundantes talament, Talameres et Birliannes thios |
| PRODUCE TO COME CHARGE BENEVA | | | SHRA | TV PROBLE | 2004TFF BILDWINGTONEY | | PRO-PART FRE | ranotovo otrus, rannumo o etembor, Tainprovo ekiloficang et Nos. |
| PRINCIPAL SI Forcelo Show-debut Randy-Pi | | Cavelo (FlayStation, Shar, etc.) | | | | Helkightpur | | Teleproverhillercompetition |
| MAGNE W Formale Manked Advanta | | PE, Courselo | | Callel Day | | | | Teresistes, laingress stillet competition, Fortunis startainment |
| OMNOSE4 19 Famely Dubi Resoluti | nor Many than 20 has y | Cascalir (PlanStotian, Bascato,) | Psychiatroper | Selon | Social Media | Melkinbour | T100-510 | Tarasisfas, Tarafarastrust |
| SYNTORDA SI Famelo Modureked Delly | T-Tilaner | Cavale Hebits | Action/Advanture, Contribution Chapter (CCC), Sale-Plant | | | Earl | | Telmannuck Befransuchting, Factorhatungingung |
| printings to Famely Shouletel Stoppin | | ACC. | Family | Frank of | | Makhiner | | Telegrapy Address action, Securidity |
| | | | | | | | | |
| SHRISHS 26 Female Dubi RandyAN | | Carrely, Habita | Artise/Advertors, First-Passen Stanter (FPS), Rate-Play | | Sanial Madia, Gaming Farrane, Game Reviews Hillage, | | | Terreferentemen, Terreninden |
| SYMPHOLOG St Family Almodulari Daily | 5-10 hours | PC, Caurelo | First Presen Shapter (FPS), Rate Playing Sense (RPG), Si | | GaningFarena, Tax Fabri Struming pletfarms (Tuit | | tipos-mos | Taraciekas, Far feedonturt sismost |
| | | | | | | | | |

Figure 5 shows the final state of the data, after changes.

2. Database Schema Design with SQL as a DDL

Schema Design

A star schema was chosen to optimize data performance and quality. The central fact table holds transaction data, while dimension tables cover customers, products, and dates. This structure supports efficient querying, reducing redundancy (as seen in figure 10).

Table Definitions

- Customers Table: Contains unique customer IDs, names, and locations (as seen in figure 6).
- Products Table: Includes product IDs, names, and category information (as seen in figure 6).
- **Transactions Table:** Holds transaction ID, customer ID (foreign key), product ID (foreign key), transaction date, and amount (as seen in figure 6).
- *Assumptions*: Data types include integers for IDs, decimals for amounts, and strings for descriptive fields. The Date from Transactions table is in DATE format to enable efficient querying.

SQL DDL and DML Commands

The sample SQL commands used to define these tables, to add some sample data can be seen in Figure 6:

```
1 -- Create tables
2 CREATE TABLE Customers (
     CustomerID INT PRIMARY KEY,
     Name VARCHAR(100),
5
     Age INT,
 6 Location VARCHAR(50)
7 );
8
9 CREATE TABLE Products (
10 ProductID INT PRIMARY KEY,
    ProductName VARCHAR(100),
11
12
     Category VARCHAR(50)
13 );
14
15 CREATE TABLE Transactions (
16 TransactionID INT PRIMARY KEY,
17 CustomerID INT.
18 ProductID INT,
    TransactionDate DATE,
19
    Amount DECIMAL(10, 2),
20
21
    FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),
22 FOREIGN KEY (ProductID) REFERENCES Products(ProductID)
23 );
24
25 -- Insert data into Customers
26 , INSERT INTO Customers (CustomerID, Name, Age, Location) VALUES
27 (1, 'John Doe', 30, 'New York'),
28 (2, 'Jane Smith', 25, 'Los Angeles'),
29 (3, 'Alice Brown', 28, 'Chicago');
30
31 -- Insert data into Products
32 ... INSERT INTO Products (ProductID, ProductName, Category) VALUES
33 (1, 'Laptop', 'Electronics'),
34 (2, 'Coffee Maker', 'Appliances'),
35 (3, 'Book', 'Literature');
37 -- Insert data into Transactions
38 VINSERT INTO Transactions (TransactionID, CustomerID, ProductID, TransactionDate, Amount) VALUES
39 (1, 1, 1, '2023-10-01', 1200.00),
40 (2, 2, 2, '2023-10-02', 150.50),
41 (3, 3, 3, '2023-10-03', 20.00);
42
43 -- Show all customers
44 SELECT * FROM Customers;
45
46 -- Show all products
47 SELECT * FROM Products;
48
49 SELECT
50 t.TransactionID,
51
    p.ProductName as Product,
     t.TransactionDate as Date,
53
54 t.Amount
55 FROM
56 Transactions t
57 JOIN
58 Customers c, Products p
59 ON
60 t.CustomerID = c.CustomerID AND t.ProductID = p.ProductID
```

Figure 6 shows the MySQL queries to create the tables: Customers, Products, Transactions, queries to add some sample data and queries to show the data from the tables

| CustomerID | Name | Age | Location |
|------------|-------------|-----|-------------|
| 1 | John Doe | 30 | New York |
| 2 | Jane Smith | 25 | Los Angeles |
| 3 | Alice Brown | 28 | Chicago |

Figure 7 shows the Customers table

| ProductID | ProductName | Category |
|-----------|--------------|-------------|
| 1 | Laptop | Electronics |
| 2 | Coffee Maker | Appliances |
| 3 | Book | Literature |

Figure 8 shows the Products table

| TransactionID | Name | Product | Date | Amount |
|---------------|-------------|--------------|------------|--------|
| 1 | John Doe | Laptop | 2023-10-01 | 1200 |
| 2 | Jane Smith | Coffee Maker | 2023-10-02 | 150.5 |
| 3 | Alice Brown | Book | 2023-10-03 | 20 |

Figure 9 shows the Transactions table

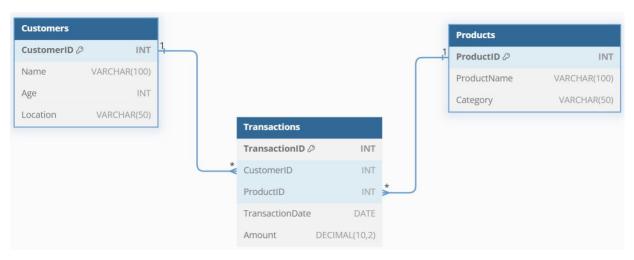


Figure 10 displays the schema creation in SQL, showing relationships among tables.

This document demonstrates a complete workflow, from data quality assessment to schema design, aimed at enhancing both data quality and performance.