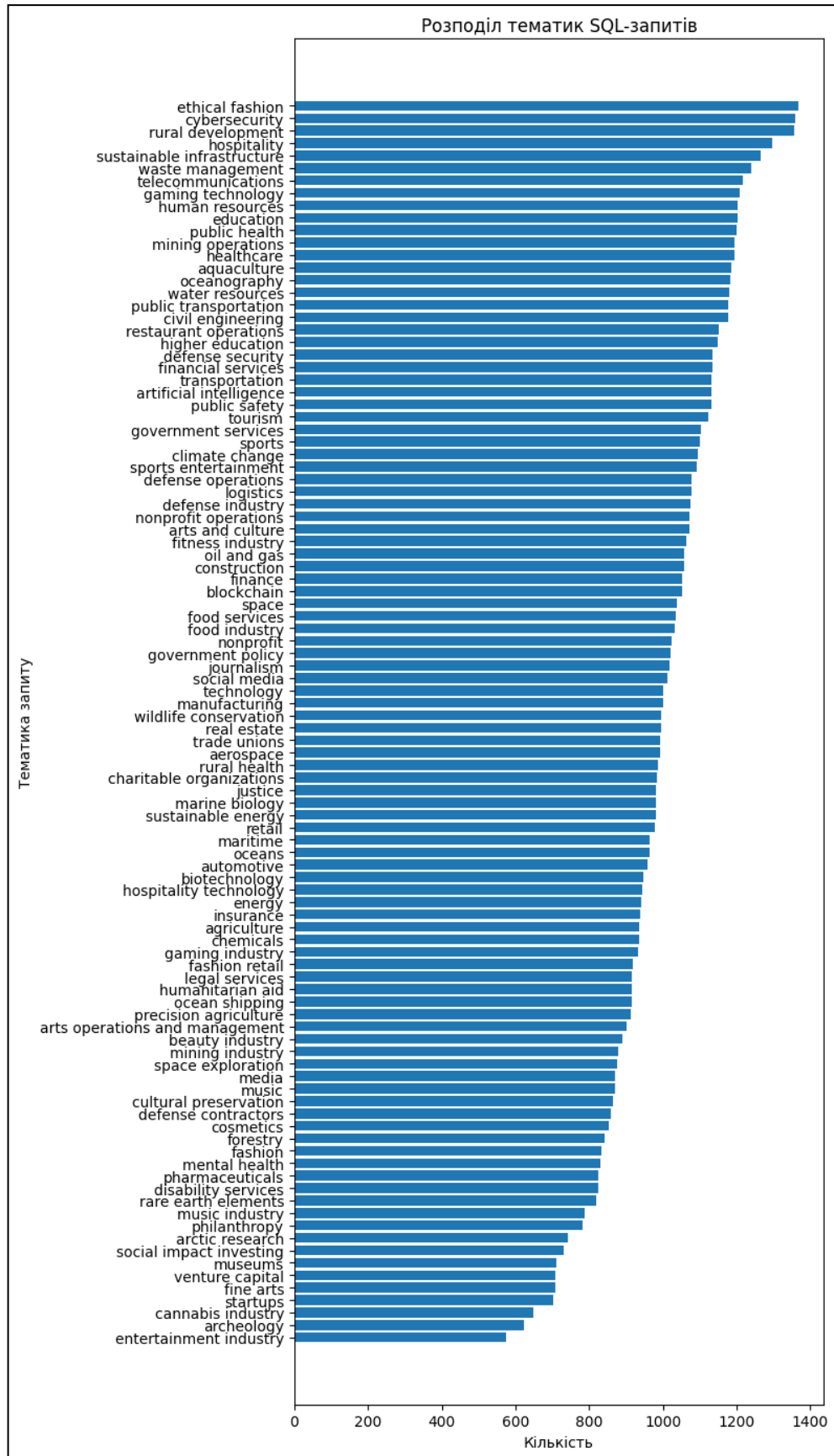


ДОДАТОК А

Діаграма розподілу тематик в тренувальному наборі



ДОДАТОК Б

Приклади входжень колонки sql_task_type

```
df[df['sql_task_type']=='analytics and reporting']['sql'].head(10).to_list()
```

```
['SELECT salesperson_id, name, SUM(volume) as total_volume FROM timber_sales JOIN salesperson ON timber_sales.salesperson_id =
id GROUP BY salesperson_id, name ORDER BY total_volume DESC;',
'SELECT equipment_type, SUM(maintenance_frequency) AS total_maintenance_frequency FROM equipment_maintenance GROUP BY equipmer
"SELECT COUNT(*) FROM marine_species WHERE location = 'Southern Ocean';",
'SELECT trader_id, stock, SUM(price * quantity) as total_trade_value, AVG(price) as avg_price FROM trade_history GROUP BY trac
'SELECT type, cost FROM (SELECT type, cost, ROW_NUMBER() OVER (ORDER BY cost DESC) as rn FROM upgrades) sub WHERE rn = 1;',
'SELECT SUM(spending) FROM defense.eu_humanitarian_assistance WHERE year BETWEEN 2019 AND 2021;',
'SELECT SpeciesName, AVG(WaterTemp) as AvgTemp FROM SpeciesWaterTemp INNER JOIN FishSpecies ON SpeciesWaterTemp.SpeciesID = Fi
E MONTH(Date) = 2 GROUP BY SpeciesName;',
'SELECT SUM(fare) FROM bus_routes WHERE route_name = 'Green Line';",
"UPDATE AVG(Property_Size) FROM Inclusive_Housing WHERE Inclusive = 'Yes';",
'SELECT d.name, AVG(f.income) FROM farmers_india f JOIN districts_india d ON f.district_id = d.id GROUP BY d.name;']
```

```
df[df['sql_task_type']=='data manipulation']['sql'].head(10).to_list()
```

```
['DELETE FROM Program_Outcomes WHERE program_id = 1002;',
'UPDATE carbon_offsets SET end_date = '2025-12-31' WHERE initiative_name = 'Initiative 1' AND country = 'Australia';",
"INSERT INTO Legislation (ID, Description, Status) VALUES (4, 'Climate Change Act', 'Draft');",
'DELETE FROM humanitarian_assistance WHERE country_id = (SELECT id FROM country WHERE name = 'Kenya') AND year = 2018;',
'INSERT INTO PolicyImpact (ID, Policy, Impact, StartDate, EndDate) VALUES (3001, 'Body Cameras for Police Officers', 'Reductio
nts by 25%', '2021-01-01', '2022-12-31');",
'INSERT INTO schedules (city, departure_time) SELECT * FROM new_schedules WHERE city = 'paris';",
'DELETE FROM donors WHERE id NOT IN (SELECT id FROM donors WHERE amount_donated > 50 AND donation_date > DATE_SUB(CURRENT_DATE
"UPDATE cybersecurity_incidents SET description = 'Unauthorized access to highly sensitive personal information' WHERE id = 1;
'UPDATE labor_statistics SET hourly_wage = hourly_wage * 1.05 WHERE job_category = 'Electricity';",
'DELETE FROM marine_life_data WHERE location = 'Mediterranean';"]
```

```
df[df['sql_task_type']=='data definition']['sql'].head(10).to_list()
```

```
['DROP TABLE vehicle_safety_testing;',
'ALTER TABLE policyholders ADD CONSTRAINT age_range CHECK (age >= 1 AND age <= 120);',
'CREATE TABLE artifacts (id INT, artifact_type VARCHAR(255), material VARCHAR(255), analysis_date DATE);',
'CREATE TABLE training_programs (id INT PRIMARY KEY, program_name VARCHAR(50), start_date DATE, end_date DATE, department VARC
(50));',
'CREATE VIEW female_fans AS SELECT * FROM fans WHERE gender = 'Female';",
'CREATE TABLE player_demographics_new AS SELECT * FROM player_demographics WHERE 1=2;',
'CREATE VIEW public_transportation_by_location AS SELECT location, GROUP_CONCAT(name) as transportation_systems FROM public_tr
cation;',
'CREATE TABLE ai_ethics (id INT PRIMARY KEY, initiative VARCHAR(50), organization VARCHAR(50), region VARCHAR(50), start_date
get FLOAT);',
'DROP TABLE redundant_billing_data;',
'CREATE TABLE infection_rates (id INT PRIMARY KEY, state VARCHAR(50), infection_rate FLOAT);']
```

```
df[df['sql_task_type']=='data retrieval']['sql'].head(10).to_list()
```

```
['SELECT * FROM defense_diplomacy WHERE participant_country IN ('China', 'India');",
"SELECT * FROM media_database WHERE (type = 'article' OR type = 'podcast') AND source = 'NPR';",
'SELECT * FROM startups WHERE founded_year = 2016;',
'SELECT * FROM older_farmers;',
'SELECT * FROM genetic_research WHERE community = 'Indigenous';",
"SELECT * FROM sustainable_practices_2 WHERE state = 'OR';",
'SELECT * FROM chemicals WHERE production_volume < 400;',
'SELECT * FROM High_Risk_Travel_Advisories WHERE country = 'Canada';",
'SELECT * FROM Machines WHERE Status != 'Operational';",
'SELECT * FROM ExcavationArtifacts;']
```

