RoomSync: A Northeastern Roommate Experience Survey and Data Synthesis

Nuri Raheem, James Lewis, Yonglin Yang
Northeastern University, Boston, MA, USA

Abstract

Despite having a co-op program that incentivizes many students to remain in student housing, Northeastern does not offer any real system for identifying students who would make for good roommates. RoomSync sought to fill this gap by collecting student data via a survey [1] and then synthesizing the results to find which questions are the best indicators of satisfaction in roommate relationships. The questions asked in the survey spanned from majors and interests to common trouble points in roommate relationships, like sleep schedule, cleanliness, and overall social interaction between the survey participant and the roommate. With 21 responses by the time the data was synthesized, we decided to supplement with some mock data for the sake of querying, which gave us enough to find relevant results from the data. From our synthesis, some of our results included that cleanliness and sleep schedule are the most critical areas that determine overall roommate satisfaction, with large discrepancies in those two areas being among the most common and most damaging conflicts in the data. Another interesting piece of data was that social interaction between roommates did not seem to be a significant predictor for roommate satisfaction, meaning that being friends and having shared interests does not guarantee a good experience. Overall, we were successfully able to collect real data from Northeastern students and used querying to synthesize what factors are most important in predicting roommate satisfaction, which could be integrated into the roommate selection process to get better matches for students.

Introduction

Our main goal in creating RoomSync was to make a complete database system that analyzes and draws conclusions from real Northeastern University roommate experiences. This is different from other roommate services that match people with their roommate simply based on their declared preferences; our database looks at past data to see what factors actually lead to good living arrangements. How students live has a big impact on how well they do in school, how happy they are, and their general feelings toward college as a whole. Most matching programs

put practical factors such as money and location above compatibility. By using real-life living situations, we can see what *actually* works.

The main features of RoomSync are:
1) An experience archive that keeps full records of 36 individuals and their past roommate situations.
2) Compatibility analysis that checks for alignment in areas like cleanliness, noise tolerance, sleep schedules, social habits, and boundary respect.
3) An interest table that contains 11 different interests and how they are shared between the respondent and their roommate.
4) Conflict documentation that records what type of conflict occurred and how that contributed to the overall experience satisfaction. This database is useful for a number of different stakeholders: university housing offices, students who are finding their roommate on their own for off-campus living, and researchers who are studying co-existence, among others.

Database Design

Overall structure and central table:
The database contains 11 interconnected tables, with roommate_experiences acting as the central hub. This table records every individual's previous roommate relationships. All other tables are linked directly to a specific experience through foreign keys(experience_id or user_id), allowing the overall analysis of what factors lead to satisfaction or conflict.

User information and lifestyle:
The users table stores name, age and major such basic demographic data. The user_lifestyle table which is linked to users records respondent's own living habits. The future_preferences table stores the expectation of future roommates for each user. This kind of design allows us to compare a user's own traits and wishes with their actual experience before.

Interest management system:
There are three tables used to track hobbies. The interests table contains all possible activities. The user_interests table creates a many-to-many link between users and their hobbies. More importantly, the shared_interests table records which specific interests were common between roommates in each past experience (experience_id), directly linking shared hobbies to relationship outcomes.

Profiles and compatibility:
Each roommate_experiences record is described by two critical satellite tables. The roommate_profiles table describes the characteristics of roommates in such experience. The

compatibility_ratings table provides quantifiable scores on how well the two individuals aligned in critical areas like cleanliness, noise, sleep, and most significantly, boundary respect.

Conflict tracking:
Conflicts are modeled to allow multiple types per experience. The conflict_types table defines standard categories. The experience_conflicts junction table then links one or more of these conflict types to a specific roommate_experiences record, telling what went through in that relationship.

Key relationships and design rationale:
The whole diagram shows a classic ”star schema” design, roommate_experiences is the fact table, surrounded by dimension tables (users, roommate_profiles) and attribute tables (compatibility_ratings, shared_interests). This structure allows powerful queries to be checked.

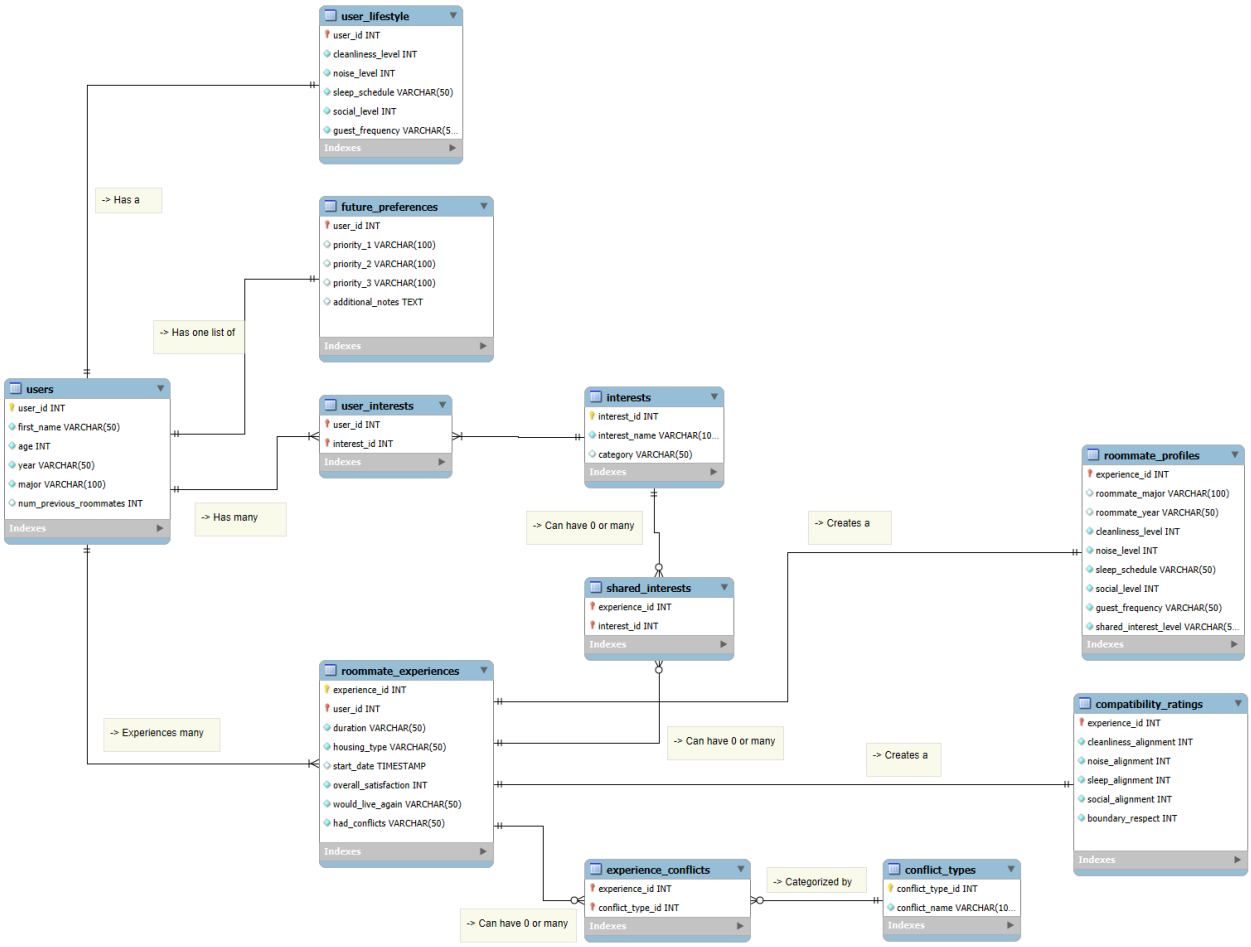

Figure 1 - MySQL Workbench database ER diagram

Data Sources and Methods

We created an in-depth 35-question survey with Google Forms [1] to gather information about responders' past roommate experiences. The survey had five sections: 1) respondent demographics, 2) respondent lifestyle characteristics, 3) their former roommate's lifestyle characteristics, 4) compatibility ratings, and 5) conflict descriptions.

This survey was sent out via many channels, such as Northeastern University Sidechat, YikYak, and Snapchat pages, as well as by asking our friends, family, and roommates. We ended up getting 21 real responses (12 at the time we built the database) from Northeastern students with different majors and living situations. While this was a good start, it was not enough for us to do a comprehensive enough database and query analysis, so we created 25 additional mock data responses for a total of 37 ‘responses’. When creating the mock data, we made sure to show similar patterns of roommate compatibility that we saw in the real data. The mock entries were made using the whole range of satisfaction levels 1-5, and show how variables are related to each other in a way that makes sense. For example, we assumed that roommates with similar sleep cycles and the same hobbies usually had higher satisfaction, so we made the responses tailored with that level in mind.

Now that all the data is in the spreadsheet [2] we began to clean the data by hand and create our INSERT statements. Our main strategies with cleaning was as writing insert statements we: 1) Standardized capitalizations, 2) Removed white spaces, 3) Expanded Abbreviations (Ex. ‘CS’ was turned into ‘Computer Science’), and 4) Separated Interest and Conflict Lists (Ex. "Sports/Fitness, Gaming, Music" turned into three independent INSERT rows). This hands-on method made sure the data was correct and let quality checks happen at every step in table creation, especially since we created the mock data ourselves as well.

User Cases and Analysis

Question 1: How does overall lifestyle alignment affect roommate satisfaction?

```
SELECT
    re.overall_satisfaction,
    COUNT(*) AS num_experiences,
    AVG(cr.cleanliness_alignment) AS cleanliness_alignment,
    AVG(cr.noise_alignment) AS noise_alignment,
    AVG(cr.sleep_alignment) AS sleep_alignment,
    AVG(cr.social_alignment) AS social_alignment,
    AVG(cr.boundary_respect) AS boundary_respect
FROM Roommate_Experiences re
JOIN Compatibility_Ratings cr ON re.experience_id = cr.experience_id
JOIN Users u ON re.user_id = u.user_id
JOIN User_Lifestyle ul ON u.user_id = ul.user_id
GROUP BY re.overall_satisfaction
ORDER BY re.overall_satisfaction DESC;
```

overall_satisfaction	num_experiences	cleanliness_alignment	noise_alignment	sleep_alignment	social_alignment	boundary_respect
5	6	4.1667	4.0000	4.3333	4.0000	4.8333
4	8	4.1250	3.7500	4.3750	4.1250	4.2500
3	4	3.2500	4.0000	2.7500	3.7500	3.2500
2	7	2.5714	2.5714	3.0000	2.0000	2.0000
1	11	2.5455	2.8182	2.6364	2.4545	1.2727

This query shows that there is a strong correlation between lifestyle alignment and roommate satisfaction on all factors that we looked at (noise, sleep, social, boundary, and cleanliness). The roommates who were the most satisfied (5/5, 6 experiences) got an average score of 4.0 or higher on cleanliness alignment (4.17), noise alignment (4.00), sleep alignment (4.33), and social habits (4.00), while the roommates who were the least happy (1/5, 11 experiences) got scores below 3.0 on all of them. The biggest difference was in boundary respect. There was a 3.56 difference between a satisfaction score of 5 and 1. This confirms to us that this is the main factor that affects how well roommates co-exist. For high satisfaction, our query emphasizes that alignment across many lifestyle aspects is needed.

Question 2: Does having more shared interests directly correlate with higher satisfaction?

```
SELECT
    rp.shared_interest_level,
    AVG(re.overall_satisfaction) AS satisfaction,
    COUNT(*) AS num_experiences,
    AVG(cr.cleanliness_alignment) AS cleanliness_alignment,
    AVG(cr.sleep_alignment) AS sleep_alignment,
    AVG(cr.social_alignment) AS social_alignment,
    AVG(cr.boundary_respect) AS boundary_respect
FROM Roommate_Experiences re
JOIN Roommate_Profiles rp ON re.experience_id = rp.experience_id
JOIN Compatibility_Ratings cr ON re.experience_id = cr.experience_id
GROUP BY rp.shared_interest_level
ORDER BY satisfaction DESC;
```

shared_interest_level	satisfaction	num_experiences	cleanliness_alignment	sleep_alignment	social_alignment	boundary_respect
Many shared interests	4.2727	11	3.9091	4.2727	4.0909	4.2727
Some shared interests	4.0000	3	4.3333	4.0000	5.0000	4.0000
Some	3.5000	2	3.5000	3.5000	3.0000	4.5000
None	1.7500	8	3.0000	3.1250	2.2500	2.1250
Few shared interests	1.6364	11	2.6364	2.7273	2.2727	1.6364
Few	1.0000	1	1.0000	1.0000	4.0000	1.0000

This question shows that shared interests are a strong indicator of how well you will get along with your roommate. For example, pairs with "many" interests had an average happiness score of 4.27/5, while pairs with "none" interests had an average satisfaction score of 1.75/5. This is a big difference of approximately 2.43 points. The query reveals that roommates who share many interests are not only happier, but they are also more compatible in every other way as well: they are more aligned in cleanliness (3.91 vs. 2.64), sleep schedules (4.27 vs. 2.73), social habits (4.09 vs. 2.27), and respect of each other's boundaries (4.27 vs. 1.64). This 2.63 point gap in boundary respect indicates that shared activities allow for better communication and mutual understanding. This creates a strong foundation for successful conflict resolution, which we found in the first query, translates to high roommate satisfaction.

Question 3: How does social interaction affect the quality of cohabitation, both between roommates and from guests invited over?

```
SELECT guest_frequency, AVG(social_level) as avg_social_level
FROM roommate_profiles
GROUP BY guest_frequency
ORDER BY avg_social_level DESC;

SELECT p.social_level, AVG(e.overall_satisfaction) as avg_satisfaction
FROM roommate_experiences as e
    JOIN roommate_profiles as p ON e.experience_id = p.experience_id
GROUP BY p.social_level
ORDER BY p.social_level DESC;

SELECT p.guest_frequency, AVG(e.overall_satisfaction) as avg_satisfaction
FROM roommate_experiences as e
    JOIN roommate_profiles as p ON e.experience_id = p.experience_id
GROUP BY p.guest_frequency
ORDER BY avg_satisfaction DESC;
```

guest_frequency	avg_social_level
Rarely/Never	3.9000
Few times a month	3.5556
Multiple times/week	3.5000
Weekly	3.4444

guest_frequency	avg_satisfaction
Rarely/Never	3.2000
Weekly	2.7778
Few times a month	2.5556
Multiple times/week	2.3750

social_level	avg_satisfaction
5	3.0769
4	1.8750
3	3.0000
2	2.5000
1	3.0000

From the data, we found the following:

- Social interaction did not appear to meaningfully change guest frequency
- Amount of guests invited by a roommate and level of social interaction between the roommates are poor predictors of overall satisfaction

While communication is an important part of any cohabitation, the data shows that general life habits like cleanliness and sleep schedule play a much larger role in roommate satisfaction than social interaction.

Question 4: What conflict types are most common and how do they affect average satisfaction?

```
SELECT t.conflict_name, COUNT(c.conflict_type_id) as num_of_conflict, AVG(e.overall_satisfaction) as avg_satisfaction_with_conflict,
    MIN(e.overall_satisfaction) as lowest_satisfaction, MAX(e.overall_satisfaction) as highest_satisfaction
FROM experience_conflicts as c
    JOIN conflict_types as t ON t.conflict_type_id = c.conflict_type_id
    JOIN roommate_experiences as e ON e.experience_id = c.experience_id
GROUP BY t.conflict_name
ORDER BY num_of_conflict DESC;
```

	conflict_name	num_of_conflict	avg_satisfaction_with_conflict	lowest_satisfaction	highest_satisfaction
▶	Cleanliness/Chores	14	1.8571	1	4
	Privacy/Boundaries	9	2.7778	1	5
	Sleep schedules	9	1.5556	1	3
	Noise	8	2.1250	1	4
	Guests	7	2.0000	1	4
	Bills/Rent	4	2.0000	1	4
	Common spaces	4	2.2500	1	3
	Food sharing	1	1.0000	1	1

From the data, we found that cleanliness was the most frequent point of contention for those surveyed, with privacy/boundaries and sleep schedules following as close seconds. This data aligns with what we found in Question 5 where high compatibility in cleanliness and privacy/boundaries resulted in the highest overall satisfaction. While whether or not a potential roommate will maintain boundaries is difficult to tell, any roommate survey should at the very least contain a section for cleanliness and preferred sleep schedule.

Question 5: How did compatibility ratings affect roommate satisfaction?

```
SELECT cr.cleanliness_alignment, AVG(e.overall_satisfaction) as avg_satisfaction, COUNT(e.overall_satisfaction) as num_experiences
FROM compatibility_ratings as cr
    JOIN roommate_experiences as e ON e.experience_id = cr.experience_id
GROUP BY cr.cleanliness_alignment
ORDER BY cr.cleanliness_alignment DESC;

SELECT cr.noise_alignment, AVG(e.overall_satisfaction) as avg_satisfaction, COUNT(e.overall_satisfaction) as num_experiences
FROM compatibility_ratings as cr
    JOIN roommate_experiences as e ON e.experience_id = cr.experience_id
GROUP BY cr.noise_alignment
ORDER BY cr.noise_alignment DESC;

SELECT cr.sleep_alignment, AVG(e.overall_satisfaction) as avg_satisfaction, COUNT(e.overall_satisfaction) as num_experiences
FROM compatibility_ratings as cr
    JOIN roommate_experiences as e ON e.experience_id = cr.experience_id
GROUP BY cr.sleep_alignment
ORDER BY cr.sleep_alignment DESC;

SELECT cr.social_alignment, AVG(e.overall_satisfaction) as avg_satisfaction, COUNT(e.overall_satisfaction) as num_experiences
FROM compatibility_ratings as cr
    JOIN roommate_experiences as e ON e.experience_id = cr.experience_id
GROUP BY cr.social_alignment
ORDER BY cr.social_alignment DESC;

SELECT cr.boundary_respect, AVG(e.overall_satisfaction) as avg_satisfaction, COUNT(e.overall_satisfaction) as num_experiences
FROM compatibility_ratings as cr
    JOIN roommate_experiences as e ON e.experience_id = cr.experience_id
GROUP BY cr.boundary_respect
ORDER BY cr.boundary_respect DESC;
```

	cleanliness_alignment	avg_satisfaction	num_experiences
▶	5	4.2500	4
	4	4.0000	11
	3	1.8571	14
	2	2.2500	4
	1	1.0000	3

	noise_alignment	avg_satisfaction	num_experiences
▶	5	3.4000	5
	4	3.9000	10
	3	2.3333	12
	2	1.6667	9

	sleep_alignment	avg_satisfaction	num_experiences
▶	5	4.1111	9
	4	3.7500	4
	3	2.2500	16
	2	1.6667	6
	1	1.0000	1

	social_alignment	avg_satisfaction	num_experiences
▶	5	4.1429	7
	4	3.8750	8
	3	1.8889	9
	2	1.4286	7
	1	2.4000	5

	boundary_respect	avg_satisfaction	num_experiences
▶	5	4.6250	8
	4	3.5714	7
	3	3.2500	4
	2	2.0000	7
	1	1.0000	10

Based on the data, compatibility ratings are extremely important when determining roommate satisfaction. All categories display clear positive linear relationships with satisfaction, and they are also additive with some categories like cleanliness and privacy having more weight than the others. Of the categories, boundaries/respect appears to have the largest positive impact on roommate satisfaction, with a boundary compatibility rating of 5 resulting in an average satisfaction of 4.63, almost a full 0.4 points higher than cleanliness, the second most impactful. Social alignment is the only category which has significant outlier data which makes sense based on the results from Question 3, however the trend is still clear.

Question 6: Does interacting more socially affect respect for boundaries/privacy?

```
SELECT
    cr.boundary_respect,
    AVG(re.overall_satisfaction) as avg_satisfaction,
    COUNT(*) as experience_count,
    AVG(cr.social_alignment) as avg_social_alignment
FROM Roommate_Experiences re
JOIN Compatibility_Ratings cr ON re.experience_id = cr.experience_id
GROUP BY cr.boundary_respect
ORDER BY cr.boundary_respect DESC;
```

	boundary_respect	avg_satisfaction	experience_count	avg_social_alignment
▶	5	4.6250	8	3.8750
	4	3.5714	7	3.8571
	3	3.2500	4	4.2500
	2	2.0000	7	2.0000
	1	1.0000	10	2.4000

From this data, we determined that while not showing a linear correlation, lower social alignment does seem to indicate a lower respect for boundaries. However, at higher boundary respect levels, no increase is seen and in fact the highest social alignment is seen at boundary respect 3, seeming to indicate that while social alignment does help build proper boundaries, it is not a sole determinant and is heavily impacted by other factors beyond social alignment.

Question 7: What are the #1 priorities among respondents for future roommates?

```
SELECT
    fp.priority_1,
    COUNT(*) as preference_count,
    AVG(re.overall_satisfaction) as avg_historical_satisfaction
FROM Future_Preferences fp
JOIN Users u ON fp.user_id = u.user_id
JOIN Roommate_Experiences re ON u.user_id = re.user_id
GROUP BY fp.priority_1
ORDER BY preference_count DESC;
```

	priority_1	preference_count	avg_historical_satisfaction
▶	Similar cleanliness	12	2.8333
	Similar social habits	4	2.2500
	Good communication	4	1.5000
	Respect for boundaries	4	3.0000
	Similar sleep schedule	4	3.0000
	Compatible noise levels	3	4.0000
	Financial responsibility	3	2.3333
	Shared interests	2	3.5000

Based on the collected data, cleanliness is by far the most important priority that respondents look for in future roommates. Cleanliness has triple the entries of the following priority which aligns with the data from Question 5 where cleanliness is the compatibility category with the second highest positive influence. An interesting piece of data from this question is that people whose highest priority is good communication have by far the lowest historical satisfaction, implying either that poor communication easily ruins roommate experiences (likely the case based on anecdotal evidence) or that people who see communication as a frequent issue themselves cause communication problems.

Conclusions

The RoomSync database analysis shows that for a roommate relationship to work, both people need to be compatible in many ways. Respecting one another's boundaries and having similar behavior in chores/cleanliness are the two most important factors that lead to good relationships. Boundary respect had the strongest impact across all 36 events. There was a 3.56-point difference in this factor between the groups with the greatest and lowest satisfaction levels, with a 0.65-point satisfaction gain for each rating point. Shared interests also appeared to be important, with a 2.43-point difference in satisfaction between "many" and "none." This factor also correlated with better alignment across all lifestyle dimensions (cleanliness, noise, sleep, and social habits), which indicates that doing things together makes it easier to talk to each other and understand each other. Our examination of individual compatibility elements confirms that elevated satisfaction (5/5) necessitates a 4.0+ alignment across all lifestyle aspects, rather than separate compatibility. Our conflict analysis found that cleanliness and housework were the most common problems (14 times, with an average satisfaction rating of 1.86). Additionally, duration patterns showed that living with the same roommate in university dorms for 12 to 18 months indicated a strong relationship (4.25/5). Our Future Preference Alignment query revealed a significant and surprising preference gap: students emphasize factors such as "similar cleanliness" (12 users) while not mentioning communication and boundary respect, which we've found yields better results as a whole. These results show that to match roommates well, matching algorithms need to be changed to focus on things like respecting boundaries, having common interests, and having a lifestyle that fits with the other person, instead of just simple things.

Author Contributions

For this project, each member contributed throughout and work was split to the best of our ability. To start, James came up with the idea of a roommate based project which was then refined as everyone gave feedback. Nuri then completed the first step of making the Google Forms which every member sent out to friends to get responses. Nuri also created the database itself while Yonglin curated the data and James built the ER diagram in MySQL Workspace. For the questions, each member handled 2-3 queries and double checked each other’s work. Work was also split for this report with each member handling at least 1 section and proof reading the document. Yonglin took the lead on the slides but everyone contributed and presented the same number of slides.

References

[1] RoomSync: Roommate Experience Student Survey. December 01, 2025. https://docs.google.com/forms/d/e/1FAIpQLSfUhWA4R42rlstZOC8w08rqz9lov1lJhKltN5AdDjVqZ4_ukQ/viewform?usp=dialog

[2] RoomSync: Roommate Experience Survey (Responses) December 01, 2025. https://docs.google.com/spreadsheets/d/1SzREmfHBEECmtoFny723ExtOr_mXPjYZ4R6sUEj63Ks/edit?usp=sharing